Radiology

ORIGINAL RESEARCH • GASTROINTESTINAL IMAGING

Colorectal Cancer: Performance and Evaluation for CT Colonography Screening— A Multicenter Cluster-randomized Controlled Trial

Anu E. Obaro, FRCR • Andrew A. Plumb, FRCR* • Steve Halligan, F Med Sci • Susan Mallett, DPhil • Paul Bassett, MSc • Paul McCoubrie, FRCR • Rachel Baldwin-Cleland, MSc • Carmen Ugarte-Cano, MSc • Phillip Lung, FRCR • Janice Muckian, DCR • Rajapandian Ilangovan, FRCR • Arun Gupta, FRCR • Charlotte Robinson, FRCR • Antony Higginson, FRCR • Ingrid Britton, FRCR • Rebecca Greenhalgh, FRCR • Uday Patel, FRCR • Evgenia Mainta, MD • Anmol Gangi, FRCR • Stuart A. Taylor, FRCR • David Burling, FRCR*

From the Centre for Medical Imaging, University College London, 43-45 Foley St, London W1W 7TS, UK (A.E.O., A.A.P., S.H., S.M., S.A.T.); Departments of Intestinal Imaging (A.E.O., R.B., C.U., P.L., J.M., R.I., A. Gupta, R.G., U.P., E.M., D.B.), St Mark's Academic Institute, St Mark's Hospital, Harrow, UK; Statsconsultancy, Amersham, UK (P.B.); Department of Radiology, Southmead Hospital, Bristol, UK (P.M.); Department of Radiology, Royal Berkshire NHS Foundation Trust, Reading, UK (C.R.); Department of Radiology, Portsmouth Hospitals University NHS Trust, Portsmouth, UK (A.H., A. Gangi); and Department of Radiology, University Hospitals of North Midlands, Stoke-on-Trent, UK (I.B.). Received June 8, 2021; revision requested July 29; revision received November 11; accepted December 13. **Address correspondence to** A.A.P. (e-mail: *andrew.plumb@ucl.ac.uk*).

Supported by 40tude Curing Colon Cancer, the Edith Murphy Foundation, the Peter Stebbings Memorial Charity, and Public Health England (funds administered by the St Mark's Hospital Foundation). A.A.P. supported by the National Institute for Health Research via the Fellowships scheme (PDF-2017-10-081). S.H. and S.A.T. supported by the UCL/UCLH Biomedical Research Centre scheme.

*A.E.O. and A.A.P. contributed equally to this work.

Conflicts of interest are listed at the end of this article.

See also the editorial by Pickhardt in this issue.

Radiology 2022; 303:361–370 • https://doi.org/10.1148/radiol.211456 • Content code: GI

Background: Most radiologists reporting CT colonography (CTC) do not undergo compulsory performance accreditation, potentially lowering diagnostic sensitivity.

Purpose: To determine whether 1-day individualized training in CTC reporting improves diagnostic sensitivity of experienced radiologists for 6-mm or larger lesions, the durability of any improvement, and any associated factors.

Materials and Methods: This prospective, multicenter cluster-randomized controlled trial was performed in National Health Service hospitals in England and Wales between April 2017 and January 2020. CTC services were cluster randomized into intervention (1-day training plus feedback) or control (no training or feedback) arms. Radiologists in the intervention arm attended a 1-day workshop focusing on CTC reporting pitfalls with individualized feedback. Radiologists in the control group received no training. Sensitivity for 6-mm or larger lesions was tested at baseline and 1, 6, and 12 months thereafter via interpretation of 10 CTC scans at each time point. The primary outcome was the mean difference in per-lesion sensitivity between arms at 1 month, analyzed using multilevel regression after adjustment for baseline sensitivity. Secondary outcomes included per-lesion sensitivity at 6- and 12-month follow-up, sensitivity for flat neoplasia, and effect of prior CTC experience.

Results: A total of 69 hospitals were randomly assigned to the intervention (31 clusters, 80 radiologists) or control (38 clusters, 59 radiologists) arm. Radiologists were experienced (median, 500–999 CTC scans interpreted) and reported CTC scans routinely (median, 151–200 scans per year). One-month sensitivity improved after intervention (66.4% [659 of 992]) compared with sensitivity in the control group (42.4% [278 of 655]; difference = 20.8%; 95% CI: 14.6, 27.0; $P < .001$). Improvements were maintained at 6 (66.4% [572 of 861] vs 50.5% [283 of 560]; difference = 13.0%; 95% CI: 7.4, 18.5; $P < .001$) and 12 (63.7% [310 of 487] vs 44.4% [187 of 421]; difference = 16.7%; 95% CI: 10.3, 23.1; $P < .001$) months. This beneficial effect applied to flat lesions (difference = 22.7%; 95% CI: 15.5, 29.9; $P < .001$) and was independent of career experience (≥1500 CTC scans: odds ratio = 1.09; 95% CI: 0.88, 1.36; $P = .22$).

Conclusion: For radiologists evaluating CT colonography studies, a 1-day training intervention yielded sustained improvement in detection of clinically relevant colorectal neoplasia, independent of previous career experience.

Clinical trial registration no. NCT02892721

Online supplemental material is available for this article.

An earlier incorrect version appeared online and in print. This article was corrected on February 28, 2022

Screening for early-stage colorectal cancer (CRC) and premalignant polyps reduces mortality by approximately 15% when testing feces for blood (the most common approach internationally) (1,2). In individuals who test positive for fecal blood (by guaiac or immunochemical methods), approximately 50% will have underlying colonic neoplasia (3). A positive fecal test result therefore requires further investigation, usually via colonoscopy, with CT colonography (CTC) used when colonoscopy is not possible (4). In CTC, multidetector CT scanners are used to produce two-dimensional and three-dimensional (3D) images of the cleansed gas-distended colon. CTC has high sensitivity in the detection of CRC and large polyps (5), and

Abbreviations

BCSP = Bowel Cancer Screening Programme, CRC = colorectal cancer, CTC = CT colonography, OR = odds ratio, 3D = three-dimensional,

Summary

One day of individualized training and feedback significantly improved radiologist sensitivity in detection of 6-mm or larger lesions at CT colonography.

Key Results

- In this prospective multicenter randomized trial, the intervention arm (80 radiologists) received 1 day of individualized training on CT colonography (CTC) image interpretation and showed sustained 16.7% improvement (63.7% [310 of 487] vs 44.4% [187 of 421]) in the detection of 6-mm or larger lesions (polyps and cancers) at 12 months ($P < .001$) compared with 59 radiologists who received no training.
- Flat lesions were more likely to be detected by radiologists assigned to the intervention arm (55.0% [458 of 832]) than by radiologists assigned to the control arm (28.5% [164 of 575], difference = 22.7%; 95% CI: 15.5, 29.9; $P < .001$).
- The positive effect of training was not associated with prior CTC career experience (≥1500 CTC scans: odds ratio = 1.09; 95% CI: 0.88, 1.36; $P = .22$).

randomized trials found no significant difference between colonoscopy and CTC for these clinically important lesions (6,7). Moreover, meta-analysis found that 3-year cancer miss rates (sometimes termed *postcolonoscopy CRC rate* for colonoscopy or *postinvestigation CRC rate* for CTC) were similar for the two tests (8,9).

In many jurisdictions, screening endoscopists undergo extensive quality assurance training to monitor key performance indicators, including adenoma detection rate and cecal intubation rate; higher adenoma detection rates are associated with lower postcolonoscopy CRC rates (10). However, radiologists are not subject to similar stipulations, and the lack of required accreditation for CTC is in sharp contrast with other screening programs using imaging. For example, accreditation and performance monitoring is mandatory in the UK National Health Service Breast Screening Programme, delivered via compulsory participation in Personal Performance in Mammographic Screening, which requires annual self-assessment and feedback on mammogram interpretation (11). Worryingly, in the English national Bowel Cancer Screening Programme (BCSP), the detection rates of CRC and high-risk colorectal polyps by CTC are only 50% of those achieved with colonoscopy (12), and missed cancers are twice as common at 3-year follow-up (13).

Our aim was to mirror the substantial improvements in mammography screening and colonoscopy achieved by assessment, training, and monitoring (14,15); thus, we hypothesized that radiologists' diagnostic accuracy for CTC screening could be improved by individualized training with ongoing feedback. To test this hypothesis, we performed a cluster-randomized trial to determine whether 1-day individualized training in CTC reporting can improve diagnostic sensitivity of experienced radiologists for 6-mm or larger lesions, the durability of any such improvement, and any variability in CTC reporting sensitivity.

Materials and Methods

Design and Participants

After approval from the University College London Research Ethics Committee (5967/003) and Health Research Authority (206876), this prospective, parallel-group, two-arm cluster-randomized superiority trial was conducted according to the publicly available protocol (16). Eligible participants were consultant radiologists or trainees within 6 months before training completion who were working in England or Wales and who routinely reported CTC results. Exclusion criteria were radiologists working outside these countries or those who could not complete 12 months of follow-up. Each hospital site constituted a cluster (Table E1 [online]). Participants provided written consent and completed a prerandomization questionnaire regarding their CTC reporting practice. Cluster randomization was performed by a statistician (P.B.) between April 2017 and September 2018 using software-generated pseudorandom numbers. Randomization was stratified by experience of the first radiologist recruited from each cluster. A cluster-randomized design was chosen because training individual radiologists at a given hospital would likely change practice among their colleagues, potentially contaminating controls. Follow-up was completed in January 2020.

Data generated or analyzed during the study may be available from the corresponding author by request.

Intervention

The intervention was an individualized training program for CTC interpretation with feedback. To assess its effect on diagnostic sensitivity, participants interpreted previously unseen CTC scans before and after randomization. Each participant interpreted 10 scans at baseline and 1, 6, and 12 months after enrollment (totaling 40 unique scans).

After baseline assessment, radiologists in the intervention arm attended a 1-day in-person workshop in which they reviewed 50 teaching scans in a 1:1 or 2:1 ratio with a CTC faculty expert (Appendix E1 [online]). Requirements for expert faculty were *(a)* career experience of more than 3000 CTC scans interpreted, *(b)* a local or national role in CTC education, or *(c)* a position at a tertiary colorectal imaging center. There was no overlap between workshop teaching cases and test scans. The intervention group received written feedback on its performance after each of the four tests with benchmarking against the rest of the sample (Appendix E2 [online]). We recorded radiologists' views regarding the training via a postworkshop survey composed of Likert scales and free text responses (Table E2 [online]). The control group did not attend the workshop or receive feedback.

CTC Tests

The test sets were chosen to represent the spectrum of luminal colorectal lesions encountered in screening practice (17). A 50%–70% prevalence of abnormal scans per set was used, similar to that after positive fecal immunochemical testing

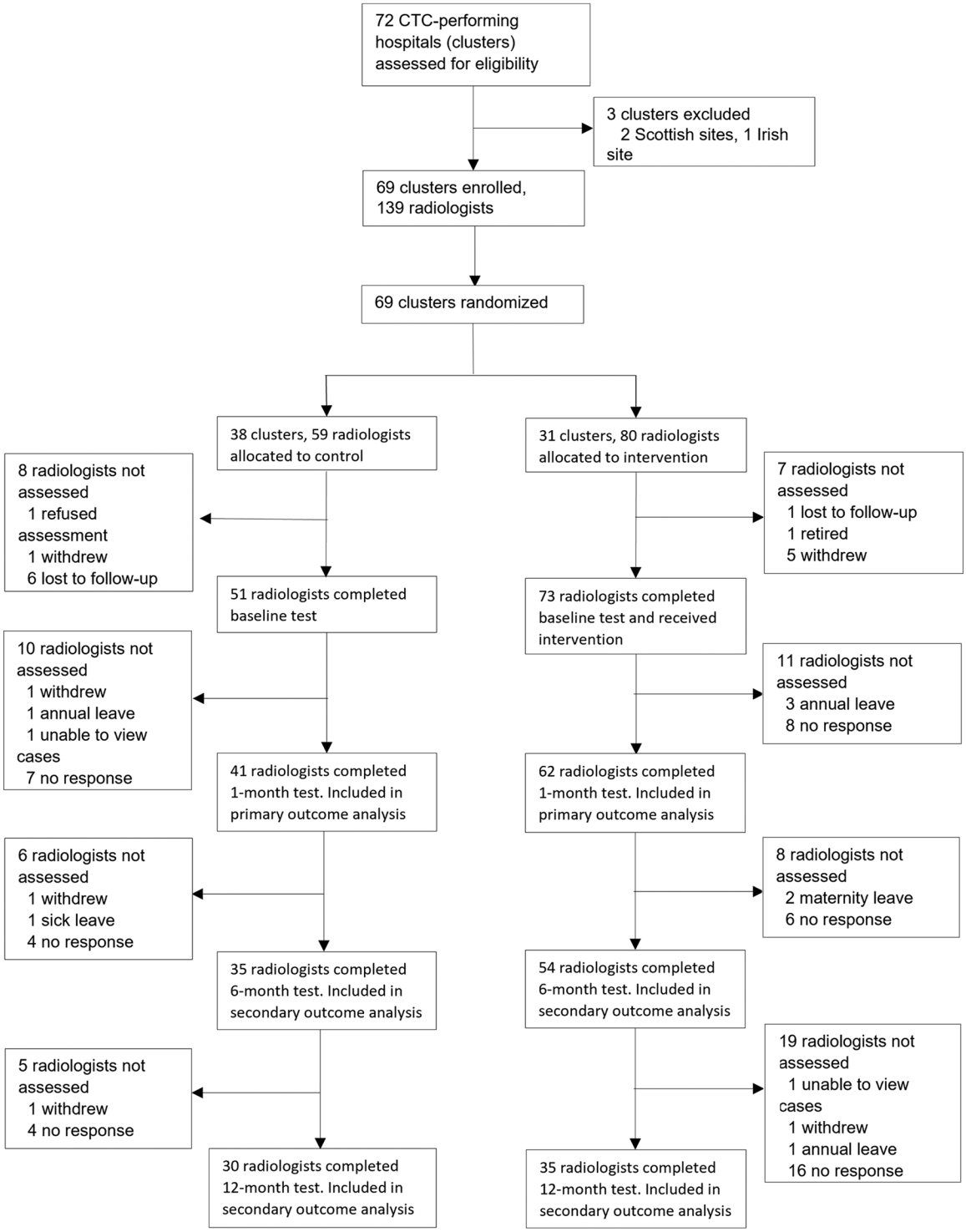

Figure 1: Flowchart of multicenter cluster-randomized controlled trial. CTC = CT colonography.

(18). Test scans were divided into true-positive results, containing 6-mm or larger lesions (polyps or cancers), and true-negative results (normal findings). True-positive results were endoscopically verified and individually scrutinized by three or more of the CTC faculty experts (A.A.P., D.B., R.I., J.M., P.L.), and consensus was reached through discussion on the size, colonic segment, conspicuity, and range of axial section numbers depicting the lesion. Scans for which consensus was not reached were excluded. Lesions of varying conspicuity were evenly distributed between all four tests. True-negative findings were scans in which both initial CTC and subsequent colonoscopy (or repeat CTC occurring 24 or more months later) showed no lesion.

We provided anonymized CTC data sets to participants on DVDs for upload and review at a CTC workstation. Participants were unaware of the prevalence of abnormalities or any clinical

information. Participants in both trial arms received the same test scans at each time point and were asked to complete an electronic case report form documenting their findings.

Outcomes

The primary outcome was the mean difference in per-lesion sensitivity (all ≥6-mm lesions) between trial arms at the 1-month time point. Secondary outcomes were the difference in per-lesion sensitivity 6 and 12 months after intervention; the difference in per-scan sensitivity 1, 6, and 12 months after intervention; the difference in per-scan specificity at all postintervention time points; and radiologist and lesion characteristics associated with higher sensitivity.

Because any given CTC scan may depict more than one polyp or cancer, scans were analyzed on a per-lesion and per-scan basis. For per-lesion analysis, an individual cancer or polyp was regarded as being detected by the interpreting radiologist if the correct section number within the prespecified range was stated along with at least two of the following parameters: *(a)* lesion type, *(b)* colonic location (within one colonic segment), and *(c)* size measurement within 50% of the reference standard (5). For per-scan analysis, we prespecified an index lesion—that is, the neoplasm with the most advanced histologic characteristics—the detection of which constituted a true-positive finding. False-positive findings were also collated via the electronic case report form.

Statistical Analyses

The sample size was based on the primary outcome: difference in 1-month sensitivity. We assumed sensitivity of 70.0% for the control arm (19) and regarded a 10.0% increase as being clinically important. Under the assumption of independent data, with a 5.0% significance level and 80.0% power, 294 lesions were required per arm. The inflation factor to account for clustering was taken as 1 + ICC (n–1) (20), where ICC is the intracluster correlation coefficient and n is the number of positive scans interpreted by each radiologist. Data from previous CTC reader studies suggested an ICC of 0.09 (21). At mean prevalence of 60.0% abnormality per test (n = 6), this inflation factor was 1.46. Therefore, we required 429 abnormal CTC scans, totaling 715 scans (429 abnormal, 286 normal) at the 1-month test, corresponding to 72 radiologists each reading 10 scans. Accounting for 10.0% dropout, we aimed to recruit 80 radiologists to each arm. A previous UK survey identified an average of three CTC-reporting radiologists per hospital (22). Therefore, we anticipated a median cluster size of two, thus aiming to recruit 40 clusters per arm.

Radiologist characteristics were compared using the χ^2 test. All outcome analyses were performed using multilevel methods and Stata, version 15.1 (Stata). Two statisticians (P.B., S.M.) used a cross-classified model, with radiologist crossed with lesion at the higher level and individual measurements at the lowest level, with separate analyses for each test time point, using logistic regression for binary outcomes (sensitivity, specificity) or Poisson regression for count outcomes (number of false-positive findings). In all analyses, the independent variables were the study arm (intervention or control) and the average outcome at baseline for each radiologist (included as a covariate). For the primary outcome, per-lesion sensitivity, cases with no lesions were omitted, and the outcome was whether the lesion was detected. In addition to the primary analyses, further analyses were performed for the same outcomes using multilevel linear regression. These were performed to obtain the absolute difference between arms, considered to be a more clinically relevant measure of size of effect. Additional prespecified analyses were performed using equivalent

Table 1: Characteristics of Radiologists Included in the Trial

Characteristic	Control Group (n = 59)	Intervention Group (n = 80)	P Value
Report for BCSP	36 (61)	59 (74)	.11
Career experience			.99
<500 CTC scans	23 (39)	31 (39)	…
500–1499 CTC scans	22 (37)	30 (38)	…
≥1500 CTC scans	14 (24)	19 (24)	…
Reporting characteristics			
3D reading performed	45 (76)	57 (71)	.51
Mean total read time spent on 3D reading (%)*	39 (5–100)	36 (5–80)	NA
Use of computer-aided detection	46 (78)	58 (72)	.46
Time to report			.34
<15 minutes	10 (17)	14 (18)	…
15–25 minutes	40 (68)	46 (58)	…
≥26 minutes	9 (15)	20 (25)	…
Previous training			.34
None	11 (19)	8 (10)	…
CTC training workshop (≤1 day)	5 (8)	8 (10)	…
CTC training workshop (≥2 days)	43 (73)	64 (80)	…
Supervised reporting	5 (8)	11 (14)	.34
Reporting with retrospective review of endoscopic findings	13 (22)	15 (19)	.70
CTC fellowship (≥3 months)	4 (7)	2 (2)	.22

Note.—Unless otherwise indicated, data in parentheses are percentages that have been rounded to the nearest whole number and may not total 100%. Some radiologists had undergone more than one form of previous CTC training, so the total is greater than 100%. BCSP = Bowel Cancer Screening Programme, CTC = CT colonography, NA = not applicable, 3D = three-dimensional.

* Data in parentheses are the range.

Table 2: Differences in Per-Lesion Detection, Per-Scan Detection, and Per-Scan Specificity between the Two Arms at Each Test Time Point

Test Time Point	Control Group (%)	Intervention Group (%)	Odds Ratio*	Difference (%)†	*P* Value
Per-lesion sensitivity					
Baseline—all‡	43.3 (420 of 969)	47.6 (661 of 1387)	1.37 (0.91, 2.05)	4.3 (−1.4, 10.0)	.13
Baseline§	43.4 (338 of 779)	49.4 (582 of 1178)	1.53 (0.99, 2.39)	6.0 (−0.2, 12.3)	.06
1 month	42.4 (278 of 655)	66.4 (659 of 992)	3.85 (2.54, 5.83)	20.8 (14.6, 27.0)	<.001
6 months	50.5 (283 of 560)	66.4 (572 of 861)	2.64 (1.79, 3.92)	13.0 (7.4, 18.5)	<.001
12 months	44.4 (187 of 421)	63.7 (310 of 487)	4.29 (2.41, 7.64)	16.7 (10.3, 23.1)	<.001
Per-scan sensitivity‖					
Baseline—all‡	29.4 (105 of 357)	35.2 (180 of 511)	1.65 (0.93, 2.94)	5.8 (−0.8, 12.4)	.09
Baseline§	28.9 (83 of 287)	37.3 (162 of 434)	2.06 (1.07, 3.95)	8.4 (1.1, 15.8)	.08
1 month	26.9 (77 of 286)	59.0 (256 of 434)	5.67 (3.40, 9.47)	29.0 (21.1, 36.9)	<.001
6 months	72.2 (177 of 245)	84.0 (316 of 376)	2.05 (1.22, 3.46)	9.9 (2.2, 17.6)	.007
12 months	59.8 (107 of 179)	75.8 (157 of 207)	8.01 (2.58, 27.1)	14.4 (8.2, 20.7)	<.001
Per-scan specificity					
Baseline—all‡	92.2 (141 of 153)	87.2 (191 of 219)	0.55 (0.23, 1.31)	−4.9 (−12.0, 2.2)	.18
Baseline§	92.7 (114 of 123)	88.2 (164 of 186)	0.55 (0.19, 1.58)	−4.5 (−12.4, 3.3)	.27
1 month	91.0 (112 of 123)	80.6 (150 of 186)	0.40 (0.19, 0.84)	−10.3 (−18.5, −2.1)	.02
6 months	90.5 (95 of 105)	84.0 (136 of 162)	0.56 (0.21, 1.51)	−5.3 (−14.5, 4.0)	.26
12 months	97.5 (117 of 120)	89.3 (125 of 140)	0.21 (0.06, 0.84)	−7.7 (−14.4, −1.0)	.03

Note.—Differing denominators are a result of the different number of lesions per test and radiologist dropout. For per-lesion sensitivity, data in parentheses are the number of lesions detected and the total number of lesions, respectively. For per-scan sensitivity, data in parentheses are the number of index lesions detected and the total number of index lesions, respectively. For per-scan specificity, data in parentheses are the number of true-negative scans correctly identified and the total number of true-negative scans, respectively. For odds ratio and difference, data in parentheses are the 95% CI.

* Odds ratio calculated as odds of detection in the intervention arm relative to the control arm, adjusted for baseline sensitivity or specificity, as appropriate.

† Percentage difference was calculated as the value in the intervention arm minus the value in the control arm, adjusted for baseline sensitivity or specificity, as appropriate.

‡ Includes data from all radiologists, including those who took no further part in the trial.

§ Data only from radiologists who provided further data.

‖ Calculated by detection of the histologically most advanced (index) lesion.

statistical methods to examine whether the intervention effect varied depending on radiologist factors (experience, BCSP reporting status, and two-dimensional vs 3D reporting) or lesion factors (size, morphologic characteristics, and segmental location). Separately, an interaction between each factor and the intervention was included in the model and, if significant, was quantified for each subgroup. $P < .05$ indicated a significant difference.

Results

Participant Characteristics

Among 72 National Health Service clusters, three were outside England or Wales and thus were excluded. Sixty-nine clusters were included, comprising 139 radiologists (134 consultants and five senior trainees; intervention arm: 31 clusters, 80 radiologists; control arm: 38 clusters, 59 radiologists) (Fig 1). Because of time constraints, recruitment closed before we reached our target sample number. Eighteen radiologists in the control arm and 18 in the intervention arm were lost to follow-up before the primary outcome. Baseline participant characteristics were well balanced between arms (Table 1). Most radiologists reported scans for the BCSP (68% [95 of 139]), median prior experience was 500–999 CTC scans, and most reported CTC scans routinely (median, 151–200 scans per year). Use of 3D visualization was similar between arms (control, 76% [45 of 59]; intervention, 71% [57 of 80]; $P = .51$), as was computer-aided detection (control, 78% [46 of 59]; intervention, 72% [58 of 80]; $P = .46$) (Table 1).

A total of 65 lesions were assessed across the four tests, comprising 12 cancers (range, 10–60 mm), five serrated lesions (range, 8–40 mm), and 48 adenomas (range, 6–50 mm); 23 lesions (35%) were flat (height <3 mm) and 46 (71%) were 10 mm or larger. At baseline testing, 1240 CTC scans were interpreted by 124 radiologists. Individual radiologist sensitivity varied widely, ranging from 15.8% (three of 19 lesions ≥6 mm detected) to 89.5% (17 of 19 lesions ≥6 mm detected), with a mean sensitivity of 45.8% (8.7 of 19 lesions) and an interquartile range of 35%–58%. Baseline per-lesion sensitivity was similar between arms (intervention, 47.6% [661 of 1387 lesions] vs control, 43.3% [420 of 969 lesions]; difference = 4.3%; 95% CI: −1.4, 10.0; $P = .13$) (Table 2) and between BCSP (48.9% [651 of 1330 lesions]) and non-BCSP (42.9% [269 of 627 lesions]) radiologists ($P = .10$) (Table E3 [online]).

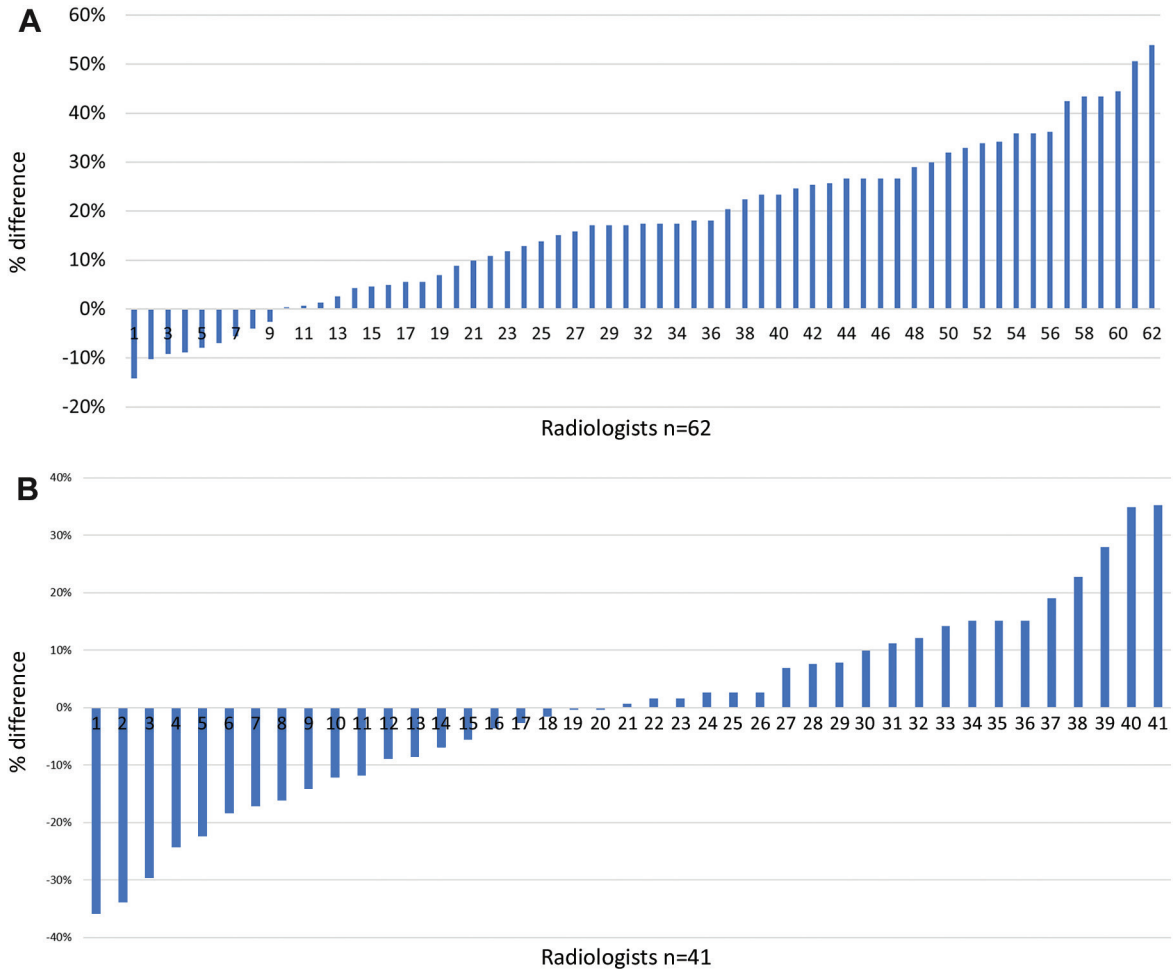

Figure 2: Differences in individual radiologist baseline and 1-month per-lesion sensitivity in detection of colorectal cancer or lesions 6 mm or larger.

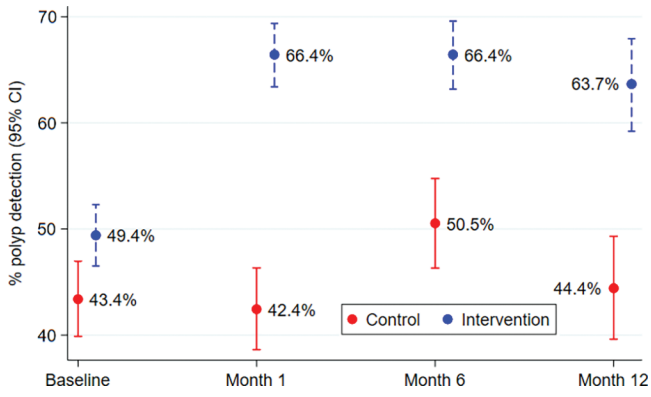

Figure 3: Per-lesion detection of colorectal cancer or lesions 6 mm or larger at each test time point for the control and intervention arms, with corresponding 95% CIs. Radiologists who dropped out before reaching the 1-month time point were omitted from the baseline calculations.

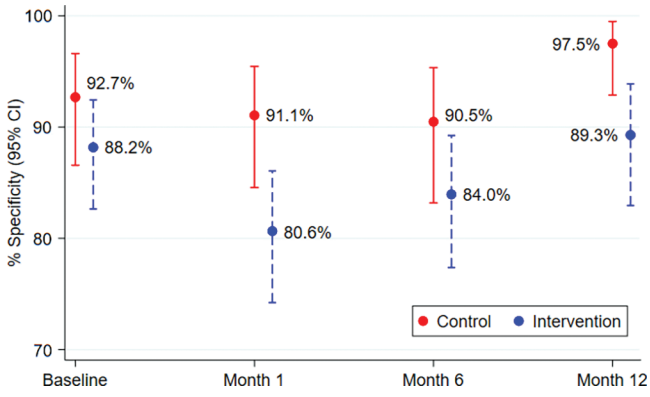

Figure 4: Per-scan specificity (denoted by identification of an index lesion) at each test time point for the control and intervention arms, with corresponding 95% CIs. Radiologists who dropped out before reaching the 1-month time point were omitted from the baseline calculations.

Outcomes

Radiologists randomly assigned to the intervention group had greater improvement in 1-month sensitivity (66.4% [659 of 992 lesions]) than did radiologists assigned to the control group (42.4% [278 of 655]; difference = 20.8%; 95% CI: 14.6, 27.0; $P < .001$) (Fig 2). This beneficial effect persisted at 6 months (intervention: 66.4% [572 of 861 lesions] vs control: 50.5% [283 of 560 lesions]; difference = 13.0%; 95% CI: 7.4, 18.5; $P < .001$) and 12 months (intervention: 63.7% [310 of 487 lesions] vs control: 44.4% [187 of 421 lesions]; difference = 16.7%; 95% CI: 10.3, 23.1; $P < .001$) (Table 2, Fig 3). Overall, the intervention arm had a 16.5% adjusted mean

Table 3: Number of False-Positive Findings at Each Test Time Point

Time Point	No. of False-Positive Findings Detected	No. of False-Positive Diagnoses per Case Interpretation*	Ratio†	Difference‡	*P* Value
Baseline—all§	…	…	1.51 (1.04, 2.20)	0.07 (0.00, 0.15)	.03
Control (*n* = 510)	65	0.13 ± 0.35	…	…	…
Intervention (*n* = 730)	147	0.20 ± 0.49	…	…	…
Baseline‖	…	…	1.83 (1.04, 3.21)	0.10 (0.04, 0.18)	.04
Control (*n* = 410)	45	0.11 ± 0.33	…	…	…
Intervention (*n* = 620)	130	0.21 ± 0.51	…	…	…
1 month	…	…	1.51 (1.04, 2.18)	0.14 (0.06, 0.21)	.03
Control (*n* = 410)	51	0.12 ± 0.43	…	…	…
Intervention (*n* = 620)	169	0.27 ± 0.59	…	…	…
6 months	…	…	0.86 (0.62, 1.20)	0.05 (-0.05, 0.14)	.37
Control (*n* = 350)	66	0.19 ± 0.50	…	…	…
Intervention (*n* = 540)	143	0.26 ± 0.69	…	…	…
12 months	…	…	2.05 (1.25, 3.36)	0.16 (0.07, 0.25)	.004
Control (*n* = 300)	24	0.08 ± 0.28	…	…	…
Intervention (*n* = 350)	96	0.27 ± 0.56	…	…	…

Note.—Data in parentheses are 95% CIs. Differing denominators are a result of the different number of lesions per test and radiologist dropout. *n* = number of CT colonography scans.

* Data are mean 6 standard deviation.

† Ratio calculated as the number of false-positive findings in the intervention arm relative to that in the control arm, adjusted for the number of false-positive findings at baseline.

‡ Difference calculated as the value in the intervention arm minus the value in the control arm, adjusted for the number of false-positive findings at baseline.

§ Includes data from all radiologists, including those who took no further part in the trial.

‖ Data only from radiologists who provided further data.

increase in per-lesion sensitivity relative to baseline during the three postrandomization test points compared with a 2.3% increase for the control arm.

When considering per-scan sensitivity for the most histologically advanced index lesion, the intervention arm showed improvement compared with the control arm at 1 month (intervention: 59.0% [256 of 434 lesions] vs control: 26.9% [77 of 286 lesions]; difference = 29.0%; 95% CI: 21.1, 36.9; $P < .001$), 6 months (intervention: 84.0% [316 of 376 lesions] vs control: 72.2% [177 of 245 lesions]; difference = 9.9%; 95% CI: 2.2, 17.6; $P = .007$), and 12 months (intervention: 75.8% [157 of 207 lesions] vs control: 59.8% [107 of 179 lesions]; difference = 14.4%; 95% CI: 8.2, 20.7; $P < .001$) (Table 2).

Per-scan specificity was similar between arms at baseline (intervention: 88.2% [164 of 186 lesions] vs control: 92.7% [114 of 123 lesions]; difference = −4.5%; 95% CI: −12.4%, 3.3%; $P = .27$). After training, per-scan specificity was lower in the intervention arm at all time points (Table 2), with the difference being significant at 1 month (intervention: 80.6% [150 of 186 lesions] vs control: 91.0% [112 of 123 lesions]; difference = −10.3%; 95% CI: −18.5%, −2.1%; $P = .02$) and 12 months (intervention: 89.3% [125 of 140 lesions] vs control: 97.5% [117 of 120 lesions]; difference = −7.7%; 95% CI: −14.4%, −1.0%; $P = .03$) (Fig 4). During all three postrandomization time points, the intervention arm had an approximately 4% adjusted mean reduction in per-scan specificity relative to baseline, compared with a 0.4% increase in the control arm (Table 2). The number of false-positive findings detected at each time point is provided in Table 3.

We found no evidence of a significant interaction between the intervention and radiologist career experience (≥1500 CTC scans: odds ratio [OR] = 1.09; 95% CI: 0.88, 1.36; $P = .22$) or lesion size (≥20 mm: OR = 2.69; 95% CI: 0.63, 11.5; $P = .35$), implying that the intervention was not dependent on either factor (Table 4). Flat lesions were more likely to be detected by radiologists assigned to the intervention group (55.0% [458 of 832 lesions]) than to the control group (28.5% [164 of 575 lesions]) (difference = 22.7%; 95% CI: 15.5, 29.9; $P < .001$), as were nonflat lesions (intervention: 71.9% [1083 of 1506 lesions] vs control: 55.0% [583 of 1059 lesions]; difference = 11.6%; 95% CI: 4.6, 18.6; $P < .001$) (Table 5). Details regarding the influence of 3D visualization, colonic segmental location, and status as a BCSP radiologist on the efficacy of the intervention are summarized in Table 5.

In both trial arms, lesion detection was highest in the rectum and lowest in the ascending and transverse colon (Table E4 [online]). In the control arm, radiologists who used 3D visualization for more than 20.0% of their interpretation time had better detection (OR = 1.48; 95% CI: 1.07, 2.05; $P < .03$), but this was not true for radiologists in the intervention arm, for which 3D interpretation had little effect on detection (OR = 0.96; 95% CI: 0.73, 1.28; $P = .33$) (Table E4 [online]). The

intervention benefitted both BCSP and non-BCSP radiologists; improvements of 15.2% (95% CI: 9.8, 20.6; $P < .001$) and 20.3% (95% CI: 12.8, 27.8; $P < .001$), respectively, were observed (Table 5).

The postworkshop survey was completed by 97% (71 of 73) of radiologists in the intervention group. Almost all either agreed (34% [24 of 71]) or strongly agreed (65% [46 of 71]) that the workshop provided useful feedback regarding their performance. All respondents agreed (27% [19 of 71]) or strongly agreed (73% [52 of 71]) that workshop cases provided additional learning opportunities, and 99% (70 of 71) would recommend the workshop to colleagues. Participation motivated 97% (69 of 71) of radiologists to improve their CTC reporting through independent study.

Table 4: Association between Radiologist or Lesion Characteristics and Study Group with Lesion-level Detection, Part 1 (All Postintervention Time Points Combined)

Study Group and Category	Detection (%)	Odds Ratio (%)	*P* Value
Both combined	…	…	.22
<500 scans	55.8 (651 of 1167)	1	…
500–1499 scans	56.8 (955 of 1681)	0.92 (0.75, 1.12)	…
≥1500 scans	60.5 (683 of 1128)	1.09 (0.88, 1.36)	…
Both combined	…	…	.35
<10 mm	44.3 (1080 of 2437)	1	…
11–19 mm	53.9 (953 of 1769)	1.89 (0.54, 6.62)	…
≥20 mm	63.0 (1336 of 2122)	2.69 (0.63, 11.5)	…

Note.— Factor times group interaction *P* value was .64 for career experience and .25 for lesion size. For detection, data in parentheses are number of lesions detected and total number of lesions, respectively. For odds ratio, data in parentheses are 95% CIs.

Discussion

Although CT colonography (CTC) is the first-choice radiologic test for colorectal cancer (CRC) screening and examination of symptomatic patients, there is no mandatory testing, accreditation, or performance monitoring for CTC reporting in most jurisdictions. This contrasts with processes for colonoscopy and other cancer screening programs. To address this deficiency, we performed a prospective, multicenter cluster-randomized controlled trial of testing, training, and feedback for experienced radiologists who routinely interpret CTC scans. In our study, a 1-day training workshop and feedback model increased radiologist sensitivity for all 6-mm or larger lesions (CRC and polyps) by 16.7%, an effect that was sustained for 12 months ($P < .001$). Improved sensitivity after intervention was observed regardless of lesion morphologic characteristics, with a 22.7% improvement in detection of flat lesions ($P < .001$). Lesion detection in the intervention arm was higher across all colonic segments ($P < .01$ for all except the descending colon [$P < .44$]) and did not depend on previous career experience (ex-

Table 5: Per-Lesion Detection according to Subgroup (All Postintervention Time Points Combined)

Subgroup	Control Group (%)	Intervention Group (%)	Odds Ratio (%)*	Difference (%)†	*P* Value
3D reporting					
<20%	43.9 (329 of 750)	68.7 (365 of 531)	4.11 (3.04, 5.55)	31.6 (20.3, 42.9)	<.001
20%–49%	48.4 (171 of 353)	67.2 (721 of 1073)	2.67 (1.95, 3.67)	17.9 (6.6, 29.2)	<.001
≥50%	46.5 (248 of 533)	61.6 (442 of 717)	2.45 (1.82, 3.28)	16.3 (5.5, 27.2)	<.001
Morphologic characteristics					
Nonflat	55.0 (583 of 1059)	71.9 (1083 of 1506)	2.46 (2.00, 3.03)	11.6 (4.6, 18.6)	<.001
Flat	28.5 (164 of 575)	55.0 (458 of 832)	4.94 (3.63, 6.71)	22.7 (15.5, 29.9)	<.001
Segment					
Cecum	36.9 (87 of 236)	58.8 (193 of 328)	2.82 (1.84, 4.30)	14.2 (2.6, 25.8)	<.001
Ascending colon	25.8 (104 of 403)	53.5 (318 of 594)	4.78 (3.35, 6.83)	25.4 (13.6, 37.3)	<.001
Transverse colon	24.6 (42 of 171)	53.4 (126 of 236)	4.64 (2.65, 8.11)	22.7 (11.4, 34.1)	<.001
Descending colon	74.3 (26 of 35)	83.3 (45 of 54)	1.53 (0.52, 4.47)	4.1 (−9.1, 17.2)	<.44
Sigmoid	56.7 (284 of 501)	74.9 (531 of 709)	2.72 (2.02, 3.66)	14.5 (4.7, 24.3)	<.001
Rectum	70.8 (204 of 288)	78.7 (328 of 417)	1.88 (1.19, 2.98)	7.3 (−3.6, 18.1)	.007
BCSP reader					
No	41.0 (260 of 634)	62.2 (377 of 606)	3.90 (2.91, 5.24)	20.3 (12.8, 27.8)	<.001
Yes	48.7 (488 of 1002)	67.1 (1164 of 1734)	2.69 (2.18, 3.31)	15.2 (9.8, 20.6)	<.001

Note.—Differing denominators are a result of the different number of lesions per test and radiologist dropout. For the control and intervention groups, data in parentheses are number of lesions detected and total number of lesions, respectively. For odds ratio and difference, data in parentheses are 95% CIs. BCSP = Bowel Cancer Screening Program, 3D = three-dimensional.

* Odds ratio calculated as odds of detection intervention group relative to the control arm, adjusted for baseline sensitivity.

† Percentage difference calculated as the value in the intervention arm minus the value in the control arm, adjusted for baseline sensitivity.

perience reading ≥1500 CTC scans: odds ratio [OR] = 1.09; 95% CI: 0.88, 1.36; P = .22) or the use of three-dimensional interpretation (OR = 0.96; 95% CI: 0.73, 1.28; P = .33). The intervention was practical to deliver, lasting only 1 day, and 99% (70 of 71) of participants would recommend it to their colleagues. Worryingly, before the training intervention, baseline sensitivity among participants was both low (45.8%) and extremely variable (range, 15.8%–89.5%) regardless of prior experience.

Previous studies on CTC reader training and testing have used novice readers (minimal prior CTC reporting experience) or fewer than 10 experienced readers rather than large representative samples of current practitioners, as we did (23–26). We recruited from 69 hospitals, representing 49.6% (69 of 139) of English CTC services (27). A prior study evaluating structured training found that approximately 175 CTC scans were required for most novice readers to achieve adequate sensitivity (24). Even so, three of nine readers did not achieve adequate performance despite prolonged training with more than 200 scans. Many professional bodies set minimum standards for CTC training by stipulating a number of studies to be reported before independent practice and documentation of annual caseload thereafter (28,29). These minimum standards are likely of limited value, as individuals achieve competence at different rates. Indeed, we found no association between career experience and lesion detection. Use of 1:1 and 2:1 training focused on individual areas of weakness and supplemented by written feedback allowed us to target learning needs to each radiologist, thereby maximizing the relevance of their training. Although the per-lesion sensitivity of radiologists after training was 66.4%, lower than previous reports from unselected screening populations (19), it is similar to the findings of another study of hard-to-detect polyps (30). Our data suggest that a model of iterated testing with subsequent individualized feedback and retraining when necessary will permit far superior sustained performance compared with accumulation of large caseloads.

Although improved detection rates might be partly offset by more false-positive referrals for colonoscopy, we found that sensitivity increased disproportionately (16.7% increased sensitivity vs 7.7% reduced specificity), meaning the net benefit would be overwhelmingly positive. This is especially relevant because patients and their doctors value sensitivity gains disproportionately over a loss of specificity (31).

This study had limitations. First, our test data set was weighted to reflect the upper end of fecal immunochemical testing prevalence and to include hard-to-detect lesions. These scans do not precisely mirror an unselected population, so caution should be applied when extrapolating our observed higher sensitivity in this test environment to other settings. Second, we closed recruitment before reaching our prespecified sample size, meaning 103 radiologists contributed to the primary outcome (initial target was 144 radiologists). However, our observed effect size was 1.6 times larger than our a priori expectation. Third, we experienced moderate loss to follow-up, albeit relatively little loss before primary end point measurement. Fourth, 21 more radiologists were randomly assigned to the intervention arm than to the control arm, despite attempts to balance this by including seven more control clusters.

In conclusion, we found that for experienced radiologists reporting CT colonography (CTC) results, a 1-day training intervention produced a sustained 16.7% improvement in the detection of clinically relevant colorectal neoplasia independent of career experience, lesion location, or morphologic characteristics. Originally, we intended to analyze pre- and posttrial lesion detection rates and positive predictive values, comparing results across arms, but the COVID-19 pandemic prevented this because of the increased workload of the recruited radiologists. This presents an avenue for future research. We believe that training and ongoing assessment should be mandated for practitioners interpreting CTC scans, and given the improvements we observed among screening radiologists, it certainly should be mandated within national screening programs. Such accreditation is already stipulated for breast cancer screening and would align CTC with colonoscopy screening. Our data suggest that radiologists would welcome this, and previous surveys have found that radiologists favor accreditation and assessment (22).

Author contributions: Guarantor of integrity of entire study, **A.A.P.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **A.E.O., A.A.P., P.M., P.L., E.M., D.B.**; clinical studies, **A.E.O., A.A.P., S.M., P.M., R.B.C., C.U.C., P.L., J.M., A. Gupta, C.R., I.B., R.G., U.P., E.M., D.B.**; statistical analysis, **A.E.O., S.H., S.M., P.B.**; and manuscript editing, **A.E.O., A.A.P., S.H., S.M., P.B., P.M., C.U.C., P.L., R.I., A. Gupta, C.R., A.H., I.B., R.G., A. Gangi, S.A.T., D.B.**

Data Sharing. Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: **A.E.O.** no relevant relationships. **A.A.P.** no relevant relationships. **S.H.** grants from the UK National Institute for Health Research; expert testimony expert testimony in medicolegal disputes; chair of the National Screening Committee, Artificial Intelligence task group. **S.M.** no relevant relationships. **P.B.** no relevant relationships. **P.M.** no relevant relationships. **R.B.** no relevant relationships. **C.U.** no relevant relationships. **P.L.** no relevant relationships. **J.M.** no relevant relationships. **R.I.** no relevant relationships. **A. Gupta** no relevant relationships. **C.R.** no relevant relationships. **A.H.** no relevant relationships. **I.B.** co-director of the National Health Service CT Colonography Training and Accreditation Programme. **R.G.** no relevant relationships. **U.P.** no relevant relationships. **E.M.** no relevant relationships. **A. Gangi** no relevant relationships. **S.A.T.** British Society of Gastrointestinal and Abdominal Radiology president; National Institute for Health Research imaging commitee chair; National Institute for Health Research Efficacy and Mechanism Evaluation Programme funding committee; share options in Motilent. **D.B.** co-director of the National Health Service CT Colonography Training and Accreditation Programme.

References

1. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348(9040):1472–1477.
2. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015;64(10):1637–1649.
3. Logan RF, Patnick J, Nickerson C, et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. Gut 2012;61(10):1439–1446.
4. Spada C, Hassan C, Bellini D, et al. Imaging alternatives to colonoscopy: CT colonography and colon capsule. European Society of Gastrointestinal Endoscopy (ESGE) and European Society of Gastrointestinal and Abdominal Radiology (ESGAR) Guideline—Update 2020. Eur Radiol 2021;31(5):2967–2982.

5. Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 2003;349(23):2191–2200.
6. Atkin W, Dadswell E, Wooldrage K, et al. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. Lancet 2013;381(9873):1194–1202.
7. Tutein Nolthenius CJ, Boellaard TN, de Haan MC, et al. Computer tomography colonography participation and yield in patients under surveillance for 6-9 mm polyps in a population-based screening trial. Eur Radiol 2016;26(8):2762–2770.
8. Obaro AE, Plumb AA, Fanshawe TR, et al. Post-imaging colorectal cancer or interval cancer rates after CT colonography: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2018;3(5):326–336.
9. Singh S, Singh PP, Murad MH, Singh H, Samadder NJ. Prevalence, risk factors, and outcomes of interval colorectal cancers: a systematic review and meta-analysis. Am J Gastroenterol 2014;109(9):1375–1389.
10. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370(14):1298–1306.
11. Gale A, Chen Y. A review of the PERFORMS scheme in breast screening. Br J Radiol 2020;93(1112):20190908.
12. Plumb AA, Halligan S, Nickerson C, et al. Use of CT colonography in the English Bowel Cancer Screening Programme. Gut 2014;63(6):964–973.
13. Derbyshire E, Plumb A, Nickerson C, et al. Post-imaging colorectal cancer in the English National Health Service Bowel Cancer Screening Programme. Gut 2019;68(Suppl 2):A189–A190.
14. Valori R. Joint Advisory Group on Gastrointestinal Endoscopy (JAG) achieves enduring large-scale change. Frontline Gastroenterol 2019;10(2):91–92.
15. Siau K, Green JT, Hawkes ND, et al. Impact of the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) on endoscopy services in the UK and beyond. Frontline Gastroenterol 2019;10(2):93–106.
16. Plumb AA, Obaro AE, Bassett P, Baldwin-Cleland R, Halligan S, Burling D. Performance and evaluation in computed tomographic colonography screening: protocol for a cluster randomised trial. medRxiv [preprint]. . Posted February 27, 2020. Accessed May 2021.
17. Plumb AA, Pathiraja F, Nickerson C, et al. Appearances of screen-detected versus symptomatic colorectal cancers at CT colonography. Eur Radiol 2016;26(12):4313–4322.
18. Moss S, Mathews C, Day TJ, et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. Gut 2017;66(9):1631–1644.
19. de Haan MC, van Gelder RE, Graser A, Bipat S, Stoker J. Diagnostic value of CT-colonography as compared to colonoscopy in an asymptomatic screening population: a meta-analysis. Eur Radiol 2011;21(8):1747–1763.
20. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. Int J Epidemiol 2015;44(3):1051–1067.
21. Halligan S, Mallett S, Altman DG, et al. Incremental benefit of computer-aided detection when used as a second and concurrent reader of CT colonographic data: multiobserver study. Radiology 2011;258(2):469–476.
22. Plumb AA, Halligan S, Taylor SA, Burling D, Nickerson C, Patnick J. CT colonography in the English Bowel Cancer Screening Programme: national survey of current practice. Clin Radiol 2013;68(5):479–487.
23. Taylor SA, Halligan S, Burling D, et al. CT colonography: effect of experience and training on reader performance. Eur Radiol 2004;14(6):1025–1033.
24. Liedenbaum MH, Bipat S, Bossuyt PM, et al. Evaluation of a standardized CT colonography training program for novice readers. Radiology 2011;258(2):477–487.
25. Fletcher JG, Chen MH, Herman BA, et al. Can radiologist training and testing ensure high performance in CT colonography? Lessons From the National CT Colonography Trial. AJR Am J Roentgenol 2010;195(1):117–125.
26. European Society of Gastrointestinal and Abdominal Radiology CT Colonography Group Investigators. Effect of directed training on reader performance for CT colonography: multicenter study. Radiology 2007;242(1):152–161.
27. Health and Social Care Information Centre. Diagnostic Imaging Dataset. https://did.hscic.gov.uk/. Published 2020. Accessed April 2021.
28. British Society of Gastrointestinal and Abdominal Radiology and The Royal College of Radiologists. Standards of practice for computed tomography colonography (CTC). https://www.rcr.ac.uk/publication/standards-practice-computed-tomography-colonography-ctc-joint-guidance-british-society. Published 2021. Accessed March 2021.
29. American College of Radiology Collaborative Committee. ACR–SAR–SCBT-MR Practice Parameter for the performance of computed tomography (CT) colonography in adults. https://www.acr.org/-/media/ACR/Files/Practice-Parameters/CT-Colonog.pdf. Published 2019. Accessed March, 2021.
30. Heresbach D, Djabbari M, Riou F, et al. Accuracy of computed tomographic colonography in a nationwide multicentre trial, and its relation to radiologist expertise. Gut 2011;60(5):658–665.
31. Boone D, Mallett S, Zhu S, et al. Patients' & healthcare professionals' values regarding true- & false-positive diagnosis when colorectal cancer screening by CT colonography: discrete choice experiment. PLoS One 2013;8(12):e80767.

Radiology

Erratum

Originally published in:
https://doi.org/10.1148/radiol.211456

Colorectal Cancer: Performance and Evaluation for CT Colonography Screening—A Multicenter Cluster-randomized Controlled Trial

Anu E. Obaro, Andrew A. Plumb, Steve Halligan, Susan Mallett, Paul Bassett, Paul McCoubrie, Rachel Baldwin-Cleland, Carmen Ugarte-Cano, Phillip Lung, Janice Muckian, Rajapandian Ilangovan, Arun Gupta, Charlotte Robinson, Antony Higginson, Ingrid Britton, Rebecca Greenhalgh, Uday Patel, Evgenia Mainta, Anmol Gangi, Stuart A. Taylor, David Burling

Erratum in:
https://doi.org/10.1148/radiol.229007

A sentence was corrected in the discussion. It now is as follows: **Fourth, 21 more radiologists were randomly assigned to the intervention arm than to the control arm, despite attempts to balance this by including seven more control clusters.**