

# Understanding and explaining variation in a national flexible sigmoidoscopy screening programme – retrospective data analyses

---

Roisin Bevan<sup>1</sup> – MClinRes

Roger G Blanks<sup>2</sup> - PhD

Claire Nickerson<sup>3</sup> - MSc

Brian P Saunders<sup>4</sup> - MD

John Stebbing<sup>5</sup> - MBBS

Richard Tighe<sup>6</sup> - MD

Andrew M Veitch<sup>7</sup> - MD

William Garrett<sup>8</sup> - MBBS

Colin J Rees<sup>9,10</sup> - MD

1. North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees
2. Cancer Epidemiology Unit, Oxford
3. NHS Cancer Screening Programmes, Sheffield
4. St Marks Hospital, London
5. Royal Surrey County Hospital NHS Foundation Trust, Guildford
6. Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich
7. The Royal Wolverhampton NHS Trust, Wolverhampton
8. Medway NHS Foundation Trust, Gillingham
9. South Tyneside NHS Foundation Trust, South Shields
10. Newcastle University, Newcastle-upon-Tyne

Corresponding author – Roisin Bevan

Correspondence address – North Tees and Hartlepool NHS Foundation Trust,  
Hardwick Rd, Stockton-on-Tees, TS19 8PE  
Correspondence email – roisinbevan@hotmail.com

### **Keywords**

**Bowel Cancer Screening, Adenoma detection, Endoscopy Quality, Key performance indicators**

**Word count – 5244 (tables/figures 892 – text body 4352)**

**Number of tables – 9**

**Number of figures – 2**

**Numbers of references - 33**

## Abstract

**Background:** A national colorectal cancer screening programme commenced in England in 2013, offering a one-off flexible sigmoidoscopy (FS) to all 55-year-olds. This programme is in addition to the biennial faecal occult blood testing programme offered to all 60-74 year olds. This study analysed data from six pilot FS screening centres to examine factors affecting the adenoma detection rate (ADR).

**Methods:** 8256 procedures performed in 55-year-old individuals were studied. Multiple regression models were used to examine the variation in ADR with withdrawal time and extent of examination, and how other factors including comfort and bowel preparation affected extent of examination.

**Findings:** The overall ADR was 9.1% and varied from 7.5-11.0% by screening centre, with a mean ADR of 11.5% in men and 6.6% in women. Multiple regression analysis showed ADR was associated with gender, withdrawal time and reported extent of examination. Extent of examination was itself associated with bowel preparation and comfort. Maximum ADR was achieved in male patients, those with good bowel preparation, deeper insertion, and where procedures were performed by endoscopists with a withdrawal time from the splenic flexure of at least 3.25 minutes in negative procedures. Mean insertion time was 4 minutes, with greater discomfort reported where insertion was not as deep.

**Interpretation:** ADR could be improved by recommending a withdrawal time from the splenic flexure of at least 3.25 minutes (ideally 3.5-4 minutes). Good bowel preparation and extent of examination were also important. Insertion time should be limited to 4 minutes as beyond this there is no additional gain but increased discomfort. The programme should maximise ADR whilst minimising discomfort; this balance is particularly important in women where both higher levels of discomfort and a lower ADR were demonstrated.

**Funding:** This project did not receive any funding.

(abstract word count 283)

## **Research in context**

### **Evidence before this study**

- Key performance indicators for colonoscopy within the BCSP are well established with high quality demonstrated
- Flexible sigmoidoscopy screening reduces colorectal cancer incidence and mortality
- Practice amongst endoscopists undertaking Flexible sigmoidoscopy screening varies and Key Performance Indicators are poorly defined

### **Added value of this study**

- Gender differences in ADR exist, but case mix adjustments do not explain endoscopist ADR variation
- ADR increases with increasing withdrawal time from the splenic flexure (up to a plateau at 4 minutes)
- Mean time taken to reach maximal insertion is 4 minutes for all segments reached, but reaching further is associated with less discomfort

### **Implications of all the available evidence**

- Key performance indicators for flexible sigmoidoscopy screening should be defined, including standards for insertion and withdrawal times, optimal depth, and bowel preparation

## BACKGROUND

Population screening for colorectal cancer (CRC) commenced in England in 2006 in the form of faecal occult blood testing (FOBT) followed by colonic investigations for FOBT positive individuals.[1] The programme aims to detect CRC at an earlier stage to improve cure rates and length of survival. A secondary aim is to detect and removes adenomas.[1] Standards are high, with adenoma detection rate (ADR) after positive FOBT of 29% in women and 43% in men.[2] An additional programme within the English NHS Bowel Cancer Screening Programme (BCSP) commenced in 2013, offering one-off flexible sigmoidoscopy (FS) to all 55-year-olds,[3] with the aim of detecting and removing adenomas from the left side of the colon, reducing CRC development via the adenoma-carcinoma sequence.[4] A large study demonstrated reduction in CRC incidence of 23% and mortality of 31% in intention-to-treat analyses,[5] forming the rationale for the new screening programme, named Bowel Scope Screening (BoSS). Six pilot sites started this programme in 2013, with roll out to the rest of the country planned by 2018. An initial benchmark ADR for endoscopists was set at 6.8%.[6] The previous trial of FS screening demonstrated an ADR of 12.1%, [5] although the cohort age-group differs from that of BoSS (55-64 years in the trial). An exploratory study to test feasibility of BoSS delivery demonstrated an ADR of 9.6%.[7] Other international FS screening programmes demonstrate ADR from 15%[8] to 18%, [9] although again the age-group varies. The FOBT arm of the BCSP has several quality indicators against which performance is monitored,[10] ensuring the programme delivers the best possible service, and supporting improvement of non-screening endoscopy services.[11] Similar quality standards are required for the BoSS programme.

This paper aims to evaluate the findings of the first 12 months of BoSS at the 6 pilot sites, with regards to ADR and the procedural factors that affect this, and to give consideration to how these results can be used to form quality standards.

## METHODS

### Study design and participants

The NHS Bowel Cancer Screening System (BCSS) uses an online database to record procedural data during screening episodes. In BoSS, data are recorded “live” during the procedure. Histological results are added later. The database was interrogated for all procedures performed at six pilot sites between May 2013 and May 2014. The pilot sites were six BCSP screening centres, which had the endoscopy unit capacity to commence the BoSS programme and represent a range of socio-economic backgrounds and geographical areas. The sites were Derby, Norwich, South of Tyne, St Mark’s (London), Surrey, West Kent and Medway, and Wolverhampton. After receiving an invitation in the post, participants self-administer an enema at home prior to attending the department for their flexible sigmoidoscopy procedure. If they are unable or unwilling to administer the enema at home, it could be given in the department on arrival.

### Procedures

Only procedures where an adenoma is removed, or biopsied, are counted as having an adenoma detected. In those procedures where one or more polyp is detected, but not removed and not biopsied (for example due to size, or total number, or due to patient factors such as anticoagulation), no adenoma is recorded, and therefore is not included in the ADR. Cases with more than 3 adenomas (either as judged clinically by the endoscopist, or proven on histology), a polyp with villous histology or high grade dysplasia, a polyp of >10mm in size, >20 hyperplastic polyps above the rectum, or where a polyp cannot be removed due to patient factors are referred on for a colonoscopy,[12] performed on a BCSP colonoscopy list by a BCSP accredited colonoscopist. Any adenomas detected at this procedure do not impact on the ADR of the original procedure. The rate of colonoscopy conversion for each site was collected.

The bowel preparation method was the same at all sites. Procedures are performed unsedated using a standard colonoscope, paediatric colonoscope, or flexible sigmoidoscope. Endoscope imagers could be used if

available/compatible with the scope in use, but their use was not mandated and not recorded onto BCSS.

Details of BoSS endoscopists were collected from information held by the BCSP or directly from the sites, and anonymised before analysis. Data analyses were restricted to those who had performed  $\geq 30$  procedures, as  $< 30$  procedures (approximately three lists) over a 12 month period was felt to reflect an inadequate number of procedures to perform analysis on. Below this level, statistical instability and potential bias could arise within the data.  $< 30$  procedures may reflect those who do not regularly perform BoSS procedures, or BoSS endoscopists at the start of their screening career when a learning curve associated with beginning screening may impact upon the results. Adenoma detection rate (ADR, the number of procedures where at least one adenoma was detected, divided by total number of procedures) was calculated for each site and each endoscopist. Insertion time was calculated by subtracting the recorded WT from the recorded total procedure time.

## Outcomes

Data were extracted from BCSS at procedural level, recording age, gender, screening site, endoscopist, date of procedure, extent of examination (as reported by the endoscopist during the procedure based on their opinion of the extent of examination reached, in conjunction with external imager guidance where available), total procedure time, withdrawal time (WT), comfort assessment, bowel preparation quality, presence of polyps/adenomas, presence of cancer, and details of complications.

## Statistical analyses

A 'gender adjusted expected to observed adenoma detection ratio' (GAADR) was calculated using the population risk of finding an adenoma in females and males (taken from the overall gender ADR in the 12 months of the programme evaluated), using this to predict how many adenomas each endoscopist should find based on their case mix, then calculating the ratio of predicted to observed adenomas.

The number of procedures where no polyp, adenoma, or cancer was detected was recorded for each site. Within the BSCP, such procedures are termed “negative procedures”. From these, several measures were derived:

Reported extent of examination for negative procedures (that is the furthest point reached during a procedure where no polyps or cancer were detected, as detection of a malignancy or several polyps that would trigger a colonoscopy referral would likely halt the procedure at an earlier point than might have been reached if no pathology was identified). A mean negative extent score was calculated by assigning the extent of examination reached a number as follows: anus(0), rectum(1), rectosigmoid(2), sigmoid(3), descending colon(4), splenic flexure(5) and transverse colon(6); expressed as a number  $\leq 6$ .

Mean negative withdrawal time (that is the time taken to withdraw the scope when no polyps or cancer were detected indicating time taken to examine the mucosa rather than being influenced by time taken to undertake polypectomy) was calculated. As WT varies depending on depth of insertion, the splenic flexure WT (SFWT) was used. As numbers of procedures reaching exactly the splenic flexure were often too small (i.e. at individual endoscopist level where few procedures were performed), the SFWT could not be accurately measured. An estimated SFWT (ESFWT) was calculated using a correction factor for the bowel segment reached in negative procedures. Withdrawal times were then grouped into broadly numerically equal-sized groups to allow for even comparison.

Univariate and multivariate analyses of factors affecting ADR were performed using binomial regression to obtain the risk ratio, 95% confidence limits for the ratio, and significance as measured by the p-value. Relative risk values were achieved by calculating the ratio of two rates. Analyses were performed using Stata (vs 14, State Corp, College Station, Texas, USA).

The quality of bowel preparation was recorded during the procedure as good, adequate, or poor (subjectively reported by the endoscopist). Validated scoring systems of bowel preparation quality for flexible sigmoidoscopy are not available. To provide a numerical measure of bowel preparation, a “bowel preparation score” for each endoscopist was produced by weighting procedures



with good preparation by 2, adequate by 1, and poor by 0 (giving a final value between 0 and 2).

### **Role of the funding source.**

No funding was received for this study. The raw data was accessed by RB, CN, RGB and CJR. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

## **RESULTS**

### **Overall outcomes**

8582 participants (4471 male, 4111 female) underwent FS in the time period studied. 8494 (99%) of 8582 were aged  $\leq 55$ . 88 participants aged  $>55$  years had “opted-in” to BoSS screening (this option to “opt in” is made available to those aged over 55 years but less than 60 and thus not yet invited to take part in the guaiac faecal occult blood testing screening programme) and were excluded from analyses due to small numbers and to avoid potential bias as their invitation method was different. 238 (2.8%) of 8494 were excluded as the procedure was performed by an endoscopist who performed less than 30 procedures in total, leaving 8256 (4299 male, 3957 female) procedures for final analyses. These procedures were performed by 44 endoscopists (mean procedure count 188 (range 32-476)). A summary of the procedures by centre is shown in table 1 demonstrating an overall ADR of 9.1% (95%CI 8.5-9.8%).

372 cases led to a referral for colonoscopy (Norwich 86 (3.1% of total), South of Tyne 53 (5.8%), St Marks 124 (5.9%), Surrey 26 (3.6%), West Kent and Medway 57 (3.5%), and Wolverhampton 44 (3.2%)). Detailed indications for and outcomes from these colonoscopies were not known.

### **Association between ADR and gender**

The ADR in females was 6.6% and in males 11.5% ( $p<0.001$ ). ADR by gender for each centre are shown in table 1. A population risk of adenomas of 0.115 (11.5%) for males and 0.067 (6.7%) for females was used to calculate a GAADR for each endoscopist. This was calculated as  $0.115 \times N_m + 0.067 \times N_F$  ( $N_m$  = number

of procedures in males,  $N_F$  = number in females). A funnel plot (Figure 1) with control lines at 95% and 99.8%, shows most endoscopists performing within the expected ranges, and one endoscopist with a notably low ADR despite a high volume of procedures.

#### Association of ADR with extent of examination and withdrawal time

The percentage of negative procedures where the splenic flexure was reached varied from 73.6% in centre F to 44.2% in centre B ( $p < 0.0001$ ) (table 2).

Negative procedures alone were considered, as those with pathology detected may have been terminated more distally in the bowel when the criteria for conversion to colonoscopy were reached (defined as detection of  $>3$  polyps felt by the endoscopist to be adenomatous, a polyp  $\geq 10$  mm in size, or detection of  $>20$  hyperplastic rectal polyps). The mean extent score varied from 4.97 in centre F to 4.35 in centre E. The lower extent score in centre E was influenced by a large number of negative procedures reaching the sigmoid colon. The negative withdrawal time from the splenic flexure (SFWT, which may be used as a measure of inspection time) varied from 2.34 to 4.76 minutes between centres ( $p < 0.0001$ ) (table 3a).

Table 3b shows the mean withdrawal time by extent of examination for negative procedures from all screening centres combined.

For individual endoscopists with fewer data we can either measure the SFWT from the procedures where the splenic flexure was reached, or make an estimate using all extent of examination data from negative procedures. Every withdrawal time from a negative procedure was multiplied by the correction factor (see table 3b) relevant to the extent of examination reached and divided by all negative procedures to obtain the ESFWT. The mean ESFWT varied between centres from 2.30 to 4.13 ( $p < 0.0001$ ). The ESFWT for the 44 endoscopists was 1.85-5.04 minutes with a mean of 3.18 minutes (figure 2).

Table 4 shows ADR against ESFWT where endoscopists are grouped into four groups by their ESFWT value. The ADR increased from 8.2% for the shortest withdrawal time group to 9.9% and 10.0% in the two highest groups (test of trend  $p = 0.021$ ). As there was little difference between the two longest

withdrawal time groups the data suggest that a SFWT of at least 3.25 minutes, and ideally 3.5-4 minutes maximises the probability of an adenoma being detected, but longer times above 4 minutes showed no evidence of increasing ADR further.

#### **Univariate and multivariate analyses of association between ADR and the explanatory factors of withdrawal time and extent of examination**

In univariate analysis (table 5) male gender demonstrated a relative risk of adenoma detection of 1.73 (95%CI 1.50-2.00), which is highly significant ( $p<0.0001$ ). The risk of adenoma detection decreased with lesser extent of examination (relative to the splenic flexure as the baseline) as would be expected, being particularly low (0.54) if only the sigmoid colon was reached. There is no evidence that reaching the transverse colon increases the detection rate above that of the splenic flexure (RR=0.89 (95%CI 0.73-1.10)). In multivariate models the results were similar.

The maximum adenoma detection rate was achieved in men where the extent of the examination reached the splenic flexure, and where WT from negative procedures was 3.5-4 minutes. Negative WT of <3.25 minutes were associated with lower detection rates. There is no evidence of further advantage from longer WT >4 minutes.

#### **Relationship between extent of examination and explanatory factors of bowel preparation and comfort**

Having shown the association between ADR and the explanatory factors of extent of examination and withdrawal time, we considered how extent of examination was influenced by bowel preparation and comfort. Table 6 demonstrates that the probability of reaching the splenic flexure depended on bowel preparation and gender. For males with good bowel preparation, 70.1% of the negative procedures reached the splenic flexure, falling to 43.0% in those with poor bowel preparation ( $p<0.0001$ ). In females these figure were lower at 56.3% and 31.4% respectively ( $p<0.0001$ ).

Table 7 shows the extent of examination was influenced by procedural discomfort, and procedures where a deeper extent of examination to the splenic

flexure or transverse colon was reached were rarely those where the patient experienced discomfort. Of 856 procedures where the transverse colon was reached, only 16 (1.9%) reported moderate or severe discomfort. By contrast, in 524 procedures where only the sigmoid colon was reached 105 (20.8%) patients had moderate or severe discomfort ( $p<0.0001$ ).

Table 8 demonstrates univariate and multivariate binomial regression models of extent of examination modelled against bowel preparation and procedure comfort showing the probability of deeper intubation to the splenic flexure declined with poor bowel preparation and increasing procedural discomfort. This suggests bowel preparation was an important factor in achieving a higher ADR. The percentage of patients with good bowel preparation varied between endoscopists from 40-80%, and between centres from 55%-71%. Between centres the mean 'bowel preparation score' varied from 1.49 to 1.69 ( $p<0.0001$ ). When considering bowel preparation score amongst endoscopists rather than sites, the mean score was 1.57 (SD 0.12, range 1.26-1.79).

## DISCUSSION

For 55-year-olds within the BoSS programme, overall ADR was 9.1%: lower than in the previous trial of FS screening,[5] similar to that of the exploratory project[7] but significantly higher than the benchmark of 6.8%. We demonstrate that there was significant difference between adenoma detection in men and women. ADR varied significantly between endoscopists, some with very low detection rates. We have demonstrated the relationship between depth of insertion and ADR, and between withdrawal time and ADR. Quality of bowel preparation was also key to ensure ADR was optimised. Each relationship is discussed further below.

The difference in ADR between men and women is expected, and demonstrated in other FS screening programmes (16.1% vs. 8.1% in the UK Flexible Sigmoidoscopy Screening Trial,[13] 22% vs. 14% (all neoplasia) in the NORCCAP study,[8] and 16.9% vs. 12.6% in the SCORE study[9]). It is also demonstrated in the BCSP FOBt screening programme (43% vs. 29%[2]). As ADR is one of the major benchmarks for performance in screening programmes, this gender

difference must be taken into account when considering quality measures. A gender-adjusted ratio could be used when analysing individual endoscopists' performance in order to ensure an accurate representation of their screening practice is assessed. Analysis of our dataset also suggests a difference in the experience of the screening procedure, with increased levels of discomfort in women (table 7), and as such a balance between comfort and pathology detection needs to be considered.

Extent of examination and withdrawal time are shown to be associated with ADR. Unlike colonoscopy screening where there is only one point of "complete examination" (caecal intubation), there is not a defined "end point" in FS screening. In BoSS, the recommendation is to examine the rectum and lower sigmoid colon and not go past the splenic flexure, with the comfort of participants during the procedure being paramount.[12] Early data on flexible sigmoidoscopy screening demonstrated improvements in CRC development after proctosigmoidoscopy with clearance of adenomas, suggesting reaching beyond the sigmoid was not necessary.[14] In the UK Flexible Sigmoidoscopy Screening Trial, a 60cm sigmoidoscope was inserted "as far as could be achieved without causing undue pain or distress (normally to the junction of the sigmoid and descending colon)".[13] The current screening programme largely uses colonoscopes inserted to the acceptable depth. Our data show no difference in insertion time between procedures reaching the sigmoid only and those reaching the splenic flexure/transverse colon, suggesting that in practice insertion time is constant regardless of extent of examination. It should be noted that the accuracy of depth of insertion is not reliable. Studies have shown correct identification of extent of examination in 50-75% of cases.[15, 16] If we make the assumption that the reported extent of examination is accurate, these data are therefore consistent with the concept that an acceptable depth is the extent of examination that is reached within a reasonable time without causing excessive discomfort. In general, "easier procedures" continue to the splenic flexure/transverse colon and more "difficult procedures" only to the sigmoid colon.

We demonstrate that women experience a much higher rate of “difficult procedures”, evidenced by longer insertion times and greater discomfort. We demonstrate higher ADR when extent of examination was greater, with optimal depth insertion to the splenic flexure. Lower ADR was demonstrated with extents of examination distal to the splenic flexure, but no additional gain if extended into the transverse colon. Interpretation of these results is complex as only a small subset of patients had an investigation that reached the transverse colon, and the reporting of extent of examination may lack accuracy. Recording of use of scope imager was not routine at the time of data collection, and more detailed analyses comparing extents of examination in those with and without a scope imager have not been performed. Extent of examination needs to be coupled with adequate time spent inspecting the mucosa on withdrawal, and is only appropriate if comfort levels are maintained. We suggest an aim of reaching the splenic flexure in order to maximise ADR; a marginally greater extent of examination than suggested by other studies.[8, 14, 17] This extent of examination should only be reached when comfort is maintained. The current data suggest that where scopes can be comfortably inserted to the splenic flexure this is not a problem. It is likely that persistent (probably unsuccessful) attempts at deep insertion cause discomfort and in these procedures extent of examination should be minimised. Table 7 and 8 show the extent of examination in negative procedures correlates to reporting of moderate/severe discomfort, showing that where such discomfort is reported the procedure is stopped at a lesser extent of examination, presumably due to the discomfort. This is reflected in BoSS guidance for endoscopists.[12] In this study, the transverse colon was reached in 856 (10.3%) procedures with the majority not reporting moderate or severe discomfort, suggesting these procedures may have reached the transverse colon precisely because there was no discomfort reported. Consideration needs to be given as to how to minimise discomfort in all procedures, and particularly in women. In women there are a lower number of adenomas to be found and the procedure is less well tolerated. One Austrian study of colonoscopic CRC screening suggested that a similar number needed to screen (NNS) to detect advanced adenomas was reached for men aged 50-54 years and women a decade

older at 60-64 years.[18] Screening women at a slightly older age (e.g. 57 or 58) might be appropriate to maximise the use of BoSS in both men and woman.

In colonoscopy practice it has been demonstrated that a WT of at least 6 minutes is associated with improved ADR,[19] and suggested that extending WT to 10 minutes in FOBt positive screening patients further improves ADR.[20] No recommendations for WT exist in flexible FS, in part due to the varying extent of examination. We have created a measure of WT, corrected for depth of insertion (the ESFWT), to assess its impact on ADR. These data would suggest a SFWT time of 3.25-4 minutes is necessary, although more work may be needed to establish if certain segments of the bowel require a longer inspection times than others, rather than an overall WT. It should be noted that there is an assumption with these data that withdrawal was of a steady speed, and that all segments of the bowel required the same pace of withdrawal to perform an accurate assessment of the mucosa. Therefore the time of 3.25 minutes to withdraw should be considered the “minimum standard”, with 3.5-4 minutes being the “aspirational standard”. The ESFWT is a tool that could be used to review individual endoscopist practice, even at relatively low numbers. In this study 24 endoscopists had WTs less than 3.25 minutes – the level below which ADR dropped significantly - suggesting scope to improve ADR by slower examination of the distal colon. By increasing WT to 4 minutes there could be a 20% increase in ADR, with an overall ADR increase of 10% – in this study, overall ADR would increase from 9.1% to 10.0%. The ESFWT has limitations; it is a novel concept, not yet tested for validity and it relies on the assumption that withdrawal is done in a smooth manner at a constant rate, and does not account for periods of time spent inspecting but not moving within the bowel. However, it provides a starting point for evaluation of withdrawal times at screening sigmoidoscopy, and warrants further work.

Quality of bowel preparation is important as it affects extent of examination reached,[21] and is likely to be one of the most important factors in detecting adenomas (as poor preparation will mean reduced ability to accurately inspect the colonic mucosa). The effect on ADR in colonoscopy is unclear, with no difference in ADR between good and fair preparation in a large US review, and a

possible reduction in proximal ADR in those with poor preparation.[22-24] UK Literature on bowel preparation in FS demonstrated improved ADR with better quality bowel preparation.[25] The same work also suggested that endoscopists with higher ADR are more likely to be critical of bowel preparation quality. We suggest a bowel preparation score is calculated for all endoscopists and centres, and a minimum standard score of 1.4 and a recommended score of at least 1.5 be introduced. Bowel preparation scores varied between centres, and may reflect list timings, preparation instructions (as the preparation was a phosphate enema at all sites, but the instructions sent out are developed by each individual site and so may vary from site to site), and should be considered for review in the future. We also acknowledge that reporting of bowel preparation is subjective, and such assessments are made by the endoscopist rather than the nurse or the SSP, and in the future, more objective measures may be more appropriate.

We use funnel plots to demonstrate individual endoscopist performance. Visual aids allowing comparison of performance are increasingly used as they allow for varying numbers of procedures, are simple to understand, and prevent the use of “league tables” to assess performance.[26-28] The plots of these data demonstrate most endoscopists performing within an acceptable range with regards to ADR, and highlight the few that may need further investigation to improve performance levels. We also used an expected to observed ratio measure (GAADR) which allows endoscopists to evaluate performance against predicted outcomes – this method of assessment has also been increasingly used to evaluate performance in varying aspects of healthcare delivery, often in terms of hospital mortality.[28-30]

We demonstrate a range of factors that impact upon ADR. Other sigmoidoscopy screening programmes and studies have demonstrated similar variations.[31-33] The UK FS screening trial[33] reported increased ADR in men, and reports other endoscopist related factors as affecting ADR, but does not examine procedure related factors such as withdrawal time or extent. It is acknowledged that this study includes a limited number of screening centres, and these centres were selected due to a willingness to participate, and a performance history and capacity to deliver the pilot. Therefore, these sites may not be completely



representative of all screening centres, and performance at FS may be heightened due to an observer effect during this pilot phase. The six sites do reflect the range of socio-economic groups in the UK and as such these findings should be generalisable at the patient level.

ADR continues to be the most widely recognised key performance indicator and as such should remain the main method for assessing quality. However it is important to look beyond headline ADR to understand what other factors may be affecting it. Understanding these factors allows for interventions to improve the quality of flexible sigmoidoscopy and thus improvement of the quality and yield of a screening programme.

#### **Suggested recommendations for the BoSS programme**

- The ADR benchmark of 6.8% should be reviewed. An aspirational ADR should be set at 9%. Funnel plots will demonstrate those performing outside the expected levels.
- GAADR can be used to ensure case-mix is taken into account when reviewing ADR.
- Endoscopists may aim to reach the splenic flexure during screening sigmoidoscopy, but within the constraints of patient comfort. Greater depth of insertion should not be at the expense of a positive patient experience.
- Withdrawal time from the splenic flexure should be not less than the minimum of 3.25 minutes, and ideally 3.5-4 minutes in negative examinations. Where necessary, ESWT can be used to ensure appropriate time is being spent on mucosal inspection.
- Insertion time should be limited to 4 minutes.
- Good quality bowel preparation should be achieved. Centres should aim for 70% of procedures to have good bowel preparation with a minimum of 60%.
- The patient experience of undergoing a BoSS procedure may affect future acceptance of other screening modalities – currently FOBt, and soon to change to FIT. The uptake of screening at the next screening round should be compared between those who have and have not undergone BoSS to assess the impact on overall screening.
- Further studies are required on a) how to minimise discomfort in women, b) the type and progression potential of adenomas detected in women, and c) whether the age at invitation for women should be increased to increase adenoma detection yield and improve the benefit/harm ratio.



## Acknowledgements

With thanks to the staff at the six pilot screening centres of St Marks, South of Tyne and Wear, West Kent and Medway, Wolverhampton, Norwich and Surrey. Special acknowledgment goes to Dr Richard Loke (West Kent and Medway Screening Centre) who passed away during the writing of this manuscript.

## Authorship

CJR and RB developed the concept for the project. RB co-ordinated the data collection, liaison with units, and writing of the manuscript. CN analysed the BCSS database, compiled the data, and contributed to analysis/writing. RB developed the statistical analysis and performed the analyses. BPS, JS, RT, AMV and WG contributed additional data from units, and contributed to development and review of the manuscript. CJR (Newcastle University) and BPS (Imperial College) hold professorships.

## Competing interests

None

## Funding

No funding was received for undertaking this work.

## References

1. Rees CJ, Bevan R. The National Health Service Bowel Cancer Screening Program: the early years. *Expert Rev Gastroenterol Hepatol*. 2013;7:421-37.
2. Logan RFA, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut*. 2012;61:1439-46.
3. Mayor S. UK committee recommends flexible sigmoidoscopy to screen for bowel cancer. *BMJ*. 2011;342:d2325.
4. Leslie A, Carey FA, Pratt NR, Steele RJC. The colorectal adenoma-carcinoma sequence. *Br J Surg*. 2002;89:845-60.
5. Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JMA, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 2010;375:1624-33.
6. NHS Cancer Screening Programmes. Letter to all Screening Centre Directors. 2015.
7. Bevan R, Rubin G, Sofianopoulou E, Patnick J, Rees CJ. Implementing a national flexible sigmoidoscopy screening program: results of the English early pilot. *Endoscopy*. 2015;47:225-31.
8. Gondal G, Grotmol T, Hofstad B, Bretthauer M, Eide TJ, Hoff G. The Norwegian Colorectal Cancer Prevention (NORCCAP) Screening Study: baseline findings and implementations for clinical work-up in age groups 50-64 years. *Scand J Gastroenterol*. 2003;38:635-42.
9. Segnan N, Senore C, Andreoni B, Aste H, Bonelli L, Crosta C, et al. Baseline findings of the Italian multicenter randomized controlled trial of "Once-Only Sigmoidoscopy" - SCORE. *J Natl Cancer Inst*. 2002;94:1763-72.
10. Lee TWJ, Rutter MD, Blanks RG, Moss SM, Goddard AF, Chilton A, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. *Gut*. 2012;61:1050-7.
11. Gavin DR, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut*. 2013;62:242-9.
12. Bowel scope screening: standard operating procedures. In: England PH, editor. <http://www.gov.uk>2018.
13. Atkin WS, Cook CF, Cuzick J, Edwards R, Northover JMA, Wardle J. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. *Lancet*. 2002;359:1291-300.
14. Atkin WS, Cuzick J, Northover JMA, Whynes DK. Prevention of colorectal cancer by once-only sigmoidoscopy. *Lancet*. 1993;341:736-40.
15. Maslekar S, Waudby P, Avery G, Monson JRT, Duthie GS. Quality assurance in flexible sigmoidoscopy: medical and nonmedical endoscopists. *Surg Endosc*. 2010;24:89-93.

16. Adam IJ, Ali Z, Shorthouse AJ. How accurate is the endoscopist's assessment of visualization of the left colon seen at flexible sigmoidoscopy? *Colorectal Dis.* 2000;2:41-4.
17. Segnan N, Patnick J, von Karsa L. European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis. 2011.
18. Ferlitsch M, Reinhart K, Pramhas S, Wiener C, Gal O, Bannert C, et al. Sex-specific prevalence of adenomas, advanced adenomas, and colorectal cancer in individuals undergoing screening colonoscopy. *JAMA.* 2011;306:1352-8.
19. Barclay RL, Vicari JJ, Doughty AS, F JJ, Greenlaw RL. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *N Engl J Med.* 2006;355:2533-41.
20. Lee TWJ, Blanks RG, Rees CJ, Wright KC, Nickerson C, Moss SM, et al. Longer mean colonoscopy withdrawal time is associate with increased adenoma detection: evidence from the Bowel Cancer Screening Programme in England. *Endoscopy.* 2013;45:20-6.
21. Froehlich F, Weitlisbach V, Gonvers J-J, Burnand B, Vader J-P. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. *Gastrointest Endosc.* 2005;61:378-84.
22. Harewood GC, Sharma VK, de Garmo P. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. *Gastrointest Endosc.* 2003;58:76-9.
23. Anderson JC, Butterly LF, Robinson CM, Goodrich M, Weiss JE. Impact of fair bowel preparation quality on adenoma and serrated polyp detection: data from the New Hampshire colonoscopy registry by using a standardized preparation-quality rating. *Gastrointest Endosc.* 2014;80:463-70.
24. Clark BT, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. *Am J Gastroenterol.* 2014;109:1714-23.
25. Thomas-Gibson S, Rogers P, Cooper S, Man R, Rutter MD, Suzuki N, et al. Judgement of the quality of bowel preparation at screening flexible sigmoidoscopy is associated with variability in adenoma detection rates. *Endoscopy.* 2006;38:456-60.
26. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Statist Med.* 2005;24:1185-202.
27. Neuburger J, Cromwell DA, Hutchings A, Black N, van der Meulen JH. Funnel plots for comparing provider performance based on patient-reported outcome measures. *BMJ Qual Saf.* 2011;20:1020-6.
28. Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg.* 2009;9:494-9.

29. Metnitz PGH, Lang T, Vesely H, Valentin A, Le Gall JR. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Med.* 2000;26:1466-72.
30. Cevasco M, Ashley SW. Quality measurement and improvement in general surgery. *Perm J.* 2011;15:48-53.
31. Pinsky PF, Schoen R, Weissfeld JL, Kramer B, Hayes RB, Yokochi L. Variability in flexible sigmoidoscopy performance among examiners in a screening trial. *Clin Gastroenterol Hepatol.* 2005;3:792.e1-.e.
32. Fracchia M, Senore C, Armaroli P, Ferraris R, Placido RD, Musso A, et al. Assessment of the multiple components of the variability in the adenoma detection rate in sigmoidoscopy screening, and lessons for training. *Endoscopy.* 2010;42:448-55.
33. Atkin W, Rogers P, Cardwell C, Cook C, Cuzick J, Wardle J, et al. Wide variation in adenoma detection rates at screening flexible sigmoidoscopy. *Gastroenterol.* 2004;126:1247-56.

**Table 1 - Endoscopists per centre, and procedural results**

Screening centre	Endoscopists with ≥30 FS (n)	Procedures			Male procedures		Female procedures	
		Total	With polyps or cancer	With adenoma (%)	n (as % of overall)	With adenoma (%)	n (as % of overall)	With adenoma (%)
A	3	824	188 (22.8)	91 (11.0)	448 (54.4)	61 (13.6)	376 (45.6)	30 (8.0)
B	9	1398	309 (22.1)	104 (7.4)	717 (51.3)	71 (9.9)	681 (48.7)	33 (4.9)
C	8	2098	420 (20.0)	202 (9.6)	1070 (51.0)	132 (12.3)	1028 (49.0)	70 (6.8)
D	11	1944	431 (22.2)	196 (10.1)	990 (50.9)	125 (12.6)	954 (49.1)	71 (7.4)
E	9	1315	230 (17.5)	99 (7.5)	707 (53.8)	66 (9.3)	608 (46.2)	33 (5.4)
F	4	677	161 (23.8)	63 (9.3)	367 (54.2)	38 (10.4)	310 (45.8)	25 (8.1)
TOTAL	44	8256	1739 (21.1)	755 (9.1)	4299 (52.1)	493 (11.5)	3957 (47.9)	262 (6.6)

**Table 2 - extent for negative procedures by centre and summary measures**

Extent	Centre A	Centre B	Centre C	Centre D	Centre E	Centre F
Anus/Rectum/Recto sigmoid	2	5	4	4	26	4
Sigmoid	47	96	47	136	179	19
Descending	188	507	447	634	326	113
Splenic	193	347	954	658	494	230
Transverse	206	134	225	81	60	150
<b>Total</b>	<b>636</b>	<b>1089</b>	<b>1677</b>	<b>1513</b>	<b>1085</b>	<b>516</b>
Mean extent score	4.87	4.47	4.80	4.44	4.35	4.97
% reaching splenic flexure	62.7	44.2	70.3	48.8	51.1	73.6

**Table 3a - Mean NWT (and relative WT compared with splenic flexure) for each screening centre by extent**

Extent	Centre A	Centre B	Centre C	Centre D	Centre E	Centre F
Anus/Rectum/Recto sigmoid	-	-	-	-	1.38 (0.59)	-
Sigmoid	1.98 (0.42)	2.03 (0.64)	1.64 (0.58)	2.60 (0.71)	1.63 (0.70)	2.74 (0.75)
Descending	3.41 (0.72)	2.72 (0.86)	2.32 (0.82)	3.21 (0.88)	1.82 (0.78)	3.08 (0.85)
Splenic	4.76 (1.00)	3.15 (1.00)	2.84 (1.00)	3.64 (1.00)	2.34 (1.00)	3.63 (1.00)
Transverse	4.50 (0.95)	3.21 (1.02)	3.04 (1.07)	4.00 (1.10)	2.15 (0.92)	4.15 (1.14)

**Table 3b - mean WT vs extent for negative procedures for all centres combined**

Extent of bowelscope	Number of Procedures	Mean withdrawal time	Correction factor*
Anus/Rectum/Rectosigmoid	45	1.42	2.23
Sigmoid	524	2.02	1.57
Descending	2215	2.72	1.17
Splenic	2876	3.17	1.00
Transverse	856	3.64	0.87

For negative procedure reaching the transverse colon a correction factor of  $3.17/3.64 = 0.87$  can be applied to all withdrawal times reaching the transverse colon to estimate the withdrawal time that would have occurred had only the splenic flexure been reached. For the sigmoid colon the correction factor is  $3.17/2.02 = 1.57$ . This correction factor enables all negative procedures to be used to calculate the estimated splenic flexure withdrawal time (ESFWT).

**Table 4 - ESFWT and ADR**

Group	ESFWT (mins) [mean]	Endoscopists (n)	Procedures (with adenomas)	ADR % (95%CI)*
1	<2.5 [2.12]	9	1877 (154)	8.2 (7.0-9.5)
2	2.5-3.24 [2.81]	15	2249 (191)	8.5 (7.4-9.7)
3	3.25-3.99 [3.53]	11	2425 (240)	9.9 (8.7-11.2)
4	≥4 [4.40]	9	1705 (170)	10.0 (8.6-11.5)

\*Test of trend  $p=0.021$

**Table 5 - Binomial regression of risk of adenoma detection by gender, estimated splenic flexure withdrawal time, and extent**

Factor	Univariate		Multivariate	
	RR (95%CI)	p-value	RR (95%CI)	p-value
<b>Gender</b>				
Female	1.00		1.00	
Male	1.73 (1.50-2.00)	<0.001	1.69 (1.46-1.95)	<0.001
<b>ESFWT</b>				
<2.5	1.00		1.00	
2.5-3.24	1.04 (0.84-1.27)	0.739	1.05 (0.86-1.29)	0.635
3.25-3.99	1.21 (0.99-1.46)	0.057	1.22 (1.00-1.48)	0.045
4.5+	1.22 (0.99-1.50)	0.066	1.22 (0.99-1.51)	0.057
<b>Extent</b>				
Transverse colon	0.89 (0.73-1.10)	0.274	0.83 (0.67-1.02)	0.069
Splenic flexure	1.00		1.00	
Descending colon	0.75 (0.64-0.88)	<0.001	0.77 (0.66-0.91)	0.001
Sigmoid colon	0.54 (0.39-0.74)	<0.001	0.59 (0.43-0.81)	0.001
Anus/rectum/rectosigmoid	0.46 (0.15-1.40)	0.173	0.48 (0.16-1.46)	0.198

**Table 6 - bowel preparation and percentage reaching splenic flexure**

Gender	Good		Adequate		Poor		Total	
	Procedures	SF(%)	Procedures	SF (%)	Procedures	SF (%)	Procedures	SF (%)
Male	1904	1335 (70.1)	1104	616 (55.8)	158	68 (43.0)	3166	2019 (63.8)
Female	2162	1218 (56.3)	981	430 (43.8)	207	65 (31.4)	3350	1713 (51.1)
Total	4066	2553 (62.8)	2085	1046 (50.2)	365	133 (36.4)	6516	3732 (57.3)



**Table 7 - Moderate or severe discomfort in negative procedures by extent (including insertion time)**

Extent	Female			Male			All		
	Procs	Mod/sev discomfort (%)	Mean Insertion time (10 <sup>th</sup> -90 <sup>th</sup> percentile)	Procs	Mod/sev discomfort (%)	Mean Insertion time (10 <sup>th</sup> -90 <sup>th</sup> percentile)	Procs	Mod/sev discomfort (%)	Mean Insertion time (10 <sup>th</sup> -90 <sup>th</sup> percentile)
Anus/Rectum/Rectosigmoid	24	7 (29.2)	1.8 (1-3)	21	3 (14.3)	2.8 (1-5)	45	10 (22.2)	2.4 (1-4)
Sigmoid	348	83 (23.9)	3.9 (2-7)	176	22 (12.5)	4.9 (2-9)	524	105 (20.8)	4.6 (2-9)
Descending	1265	145 (11.5)	4.0 (2-7)	950	58 (6.1)	4.7 (2-8)	2215	203 (9.2)	4.4 (2-8)
Splenic	1413	109 (7.7)	4.0 (2-6)	1463	47 (3.2)	4.7 (2-8)	2876	156 (5.4)	4.3 (2-7)
Transverse	300	11 (3.7)	4.2 (2-6)	556	5 (0.9)	5.2 (2-9)	856	16 (1.9)	4.5 (2-7)
All	3350	355 (10.6)		3166	135 (4.3)		6516	490 (7.5)	

**Table 8 - Binomial regression of risk of reaching splenic flexure for negative procedures by gender, bowel prep and comfort**

Factor	Univariate		Multivariate	
	RR (95%CI)	p-value	RR (95%CI)	p-value
<b>Gender</b>				
Male	1.00		1.00	
Female	0.80 (0.76-0.84)	<0.001	0.83 (0.80-0.87)	<0.001
<b>Bowel prep</b>				
Good	1.00		1.00	
Adequate	0.80 (0.76-0.84)	<0.001	0.79 (0.76-0.83)	<0.001
Poor	0.58 (0.51-0.67)	<0.001	0.58 (0.51-0.67)	<0.001
<b>Discomfort</b>				
None	1.00		1.00	
Minimal	0.92 (0.88-0.96)	<0.001	0.96 (0.92-1.00)	0.055
Mild	0.77 (0.72-0.83)	<0.001	0.82 (0.76-0.88)	<0.001
Mod/severe	0.55 (0.49-0.63)	<0.001	0.58 (0.51-0.66)	<0.001

