

# National audit of the sensitivity of double-contrast barium enema for colorectal carcinoma, using control charts

For the Royal College of Radiologists Clinical Radiology Audit Sub-Committee

D.J. Tawn<sup>a,\*</sup>, C.J. Squire<sup>a</sup>, M.A. Mohammed<sup>b</sup>, E.J. Adam<sup>a</sup>

<sup>a</sup>Clinical Radiology Audit Sub-Committee, Royal College of Radiologists, 38 Portland Place, London, and  
<sup>b</sup>Department of Public Health and Epidemiology, University of Birmingham, Birmingham, UK

Received 26 January 2004; received in revised form 2 August 2004; accepted 19 September 2004

## KEYWORDS

Barium enema examination; Control charts; Colorectal neoplasms; National audit

**AIM:** To audit the sensitivity of double-contrast barium enema (DCBE) for colorectal carcinoma, as currently practised in UK departments of radiology.

**METHODS:** As part of its programme of national audits, the Royal College of Radiologists Clinical Radiology Audit Sub-Committee undertook a retrospective audit of the sensitivity of DCBE for colorectal carcinoma during 2002. The following targets were set: demonstration of a lesion  $\geq 95\%$ ; correct identification as a carcinoma  $\geq 90\%$ .

**RESULTS:** Across the UK, 131 departments took part in the audit, involving 5454 examinations. The mean demonstration rate was 92.9% and the diagnosis rate was 85.9%, slightly below the targets set. The equivocal rate (lesion demonstrated, but not defined as malignant) was 6.9%, the perception failure rate was 2.8% and the technical failure rate was 4.4%. Control-chart methodology was used to analyze the data and to identify any departments whose performance was consistent with special-cause variation.

**CONCLUSION:** When compared with the diagnosis rate (84.6%) and demonstration rate (92.7%) reported in the Wessex Audit 1995, [Thomas RD, Fairhurst JJ, Frost RA. Wessex regional audit: barium enema in colo-rectal carcinoma. *Clin Radiol* 1995;50:647-50.<sup>1</sup>] a similar level of performance was observed in the NHS today, implying that the basic process for undertaking and reporting DCBE has remained relatively unchanged over the last few years. Improvement in the future will require fundamental changes to the process of reporting DCBE, in order to minimize the perception failure rate and accurately to describe lesions, so reducing the equivocal rate. Control-chart methodology has a useful role in identifying strategies to deliver continual improvement.

© 2005 The Royal College of Radiologists. Published by Elsevier Ltd. All rights reserved.

## Introduction

Double-contrast barium enema (DCBE) is a standard technique for investigating colonic disease, and is widely used in the diagnosis of colorectal cancer.<sup>1-6</sup> Ideally, 100% of such lesions would be demonstrated on DCBE and diagnosed (interpreted) correctly. Unfortunately, 100% rates are seldom, if ever, attainable<sup>2-6</sup> because some lesions cannot

\* Guarantor and correspondent: D.J. Tawn, Department of Radiology, Royal Bournemouth Hospital, Castle Lane East, Bournemouth BH7 7DW, UK. Tel.: +44 1202 704577; fax: +44 1202 704035.

E-mail address: julian.tawn@rbch-tr.swest.nhs.uk (D.J. Tawn).

be seen, because it is sometimes not possible to be sure that the lesion is malignant, and because human error in perceiving and reporting the lesion cannot be ruled out. Less than 100% performance can have potentially adverse consequences for patient outcomes and satisfaction, stemming from delayed diagnosis. In 1995, the Wessex Regional Audit<sup>1</sup> found the average demonstration rate of 10 radiology departments was 92.7% (96.8% after correction for errors of perception) and the diagnosis rate was 84.6%. The authors proposed targets of 97% demonstration rates and 94% diagnosis rates.

We undertook a national audit of DCBE demonstration and diagnosis rates to determine if any improvement had occurred since the Wessex Regional Audit (1995), and also to provide guidance on how current performance could be improved. For the latter purpose, we adopted Shewhart's theory of variation,<sup>7,8</sup> which seeks to provide guidance on the action required to deliver continual improvement. After some debate, we adopted the targets set out in Table 1. These targets are higher than those reported by the Wessex Audit and lower than those it proposed for re-audit, but were considered by the Audit Sub-Committee as attainable procedural and personal standards.

## Methods

All 302 UK NHS radiology departments on the Royal College of Radiology (RCR) audit database were, through their nominated contact, invited to participate; 131 did so. A consultant radiologist from each department was asked to undertake this retrospective study and to obtain, from the pathology department, a list of colorectal cancer cases with a histologically confirmed diagnosis recorded during the year 2001. This list was then used to identify all such patients who had had a DCBE at any time during the previous year. The period of 1 year was chosen to allow direct comparison with the Wessex Audit. For each such patient identified, the DCBE films and associated report were compared with the histology report.

Contextual information was also sought. Departments were asked how many consultants, trainees and radiographers were performing barium enemas, and to add any comments they wished. The results of comparisons were classified using the scheme set out in Table 2.

Departments were asked to review the films of all cases where there was an equivocal radiology report, or a mismatch between the radiology and histology reports. We recommended that departments undertook their review of these cases at a radiology discrepancy meeting. Equivocal cases were, on retrospective review, classified as clearly, probably, or not malignant. In cases where there was no match between the radiology report and histology report, we asked to know whether a lesion was visible on retrospective review, and whether further investigations had been recommended and why.

The diagnosis rate (target  $\geq 90\%$ ) was obtained from the match score; the demonstration rate (target  $\geq 95\%$ ) was the match score plus the equivocal score. Each department returned its results on a summary data collection form to the RCR Audit Officer for data entry and analysis. Participating departments were sent a copy of the results, in a format that allowed them to identify themselves and compare their performance with that of other UK departments, while maintaining the anonymity of those departments.

## Shewhart's theory of variation and control charts

Our method for guiding improvement action was based on Shewhart's theory of variation,<sup>7,8</sup> which classifies variation according to the action required in order to reduce it. Shewhart identified two types of variation: common cause and special cause. Common-cause variation is expected variation attributable to "chance", which is inherent in every process and affects every operator. To reduce common-cause variation, we need to change the process. By contrast, special-cause variation is exceptional variation not attributable to "chance". It arises from special circumstances affecting only some operators. Special-cause variation can

**Table 1** Targets for the national audit of results of double-contrast barium enema in cases of proven colorectal carcinoma

Result	Target
Demonstration	For cases with a barium enema performed within 1 year of the diagnosis, the lesion should be demonstrated in at least 95% of cases
Diagnosis	The lesion should be correctly identified as a carcinoma in at least 90% of cases
Further investigation recommended	For cases where a lesion is shown but not diagnosed as a carcinoma (equivocal cases), further investigation should be recommended in 100% of cases

**Table 2** Classification scheme of results of double-contrast barium enema

Result	Definition
Diagnosis (=match)	The diagnoses from the DCBE and histology reports the same, i.e. both mention colorectal carcinoma or its synonym, at the same site: reports match
Equivocal Demonstration	A lesion reported, but carcinoma not mentioned A lesion demonstrated, whether or not reported as possible carcinoma (= diagnosis rate + equivocal rate)
Perceptive mismatch	A lesion apparent on review of the DCBE but not seen (or reported) originally
Technical mismatch	A lesion not visible even in retrospect. (includes incomplete double-contrast barium enema)

produce exceptionally good or bad outcomes. To reduce unfavourable special-cause variation, we need to identify the special cause and act on it. In the case of favourable special-cause variation, we likewise need to investigate and learn from it.

This variation is usefully displayed graphically on a control chart. These charts have three lines, i.e. a central line showing the mean, and upper and lower lines set at three standard deviations above and below the central line, respectively.<sup>8,9</sup> The criterion of three standard deviations from the central line has been used to minimize the "false alarm" rate, and to emphasize the need to identify special causes of variation; it is economical to find and eliminate these. The control limits are intended as guidelines, which help the user differentiate between common-cause and special-cause variation. Points on or above the lower control limit, but below the upper control limit, arise from common-cause variation about the mean. Points on or above the upper control limit, and those below the lower control limit, arise from special-cause variation; this warrants further investigation to identify the cause and, where appropriate, to eliminate it.

For each aspect of the performance of DCBE under consideration, results are presented as a funnel plot, i.e. as a control chart, with limits, but plotting the percentage score against sample size. This produces a funnel effect, showing how sample size impacts on the variability of results.<sup>10,11</sup> We calculated the control limits using Spiegelhalter's approach,<sup>10</sup> which is an exact method of calculating the limits. The discreteness of the underlying binomial distribution of the data is apparent in the resulting "saw-tooth" effect in the control limits. We elected not to smooth these curves, as this would undermine the purpose of the exact calculation.

## Results

**Fig. 1** displays on a control chart the diagnosis rate for all departments. The overall diagnosis rate,

horizontal line, was 85.9% (4687/5454), range 50% to 100%. The control chart divides the departments into three distinct groups. The 7 departments on or above the upper control limit indicate that special-cause variation rather than chance has produced these high diagnosis rates. The 3 departments below the lower control limit again imply special-cause variation, this time producing low diagnosis rates, which for these sample sizes are extremely unlikely to have arisen by chance alone. The scores of the remaining 121 departments appearing on or above the lower control limit, but below the upper limit, imply common-cause variation.

**Fig. 2** displays on a control chart the equivocal rate for all departments. The overall equivocal rate was 6.9% (379/5454), range 0 to 35.5%. The 1 department below the lower control limit indicates special-cause variation, producing especially low equivocal rates. The 10 departments on or above the upper control limit imply special-cause variation, producing high equivocal rates that are unlikely to have arisen by chance alone. The remaining 120 departments with scores on or above the lower control limit and below the upper limit imply common-cause variation only.

**Fig. 3** displays on a control chart the lesion demonstration rate (diagnosis rate plus equivocal rate) for all departments. The overall demonstration rate, horizontal line, was 92.9% (5066/5454), range 60% to 100%. In all, 11 departments have scores on or above the upper control limit, implying that these high demonstration rates for these sample sizes are very unlikely to have arisen by chance alone, and indicating special-cause variation. There is no department with a score below the lower control limit, and the remaining 122 departments appearing on or above the lower control limit, but below the upper limit, imply common-cause variation.

We also plotted control charts for perception failure and technical failure separately. **Fig. 4** displays on a control chart the perception failure rate for all departments. The overall perception failure rate was 2.8% (150/5454), range 0 to 21.4%. The 2 departments on or above the upper control

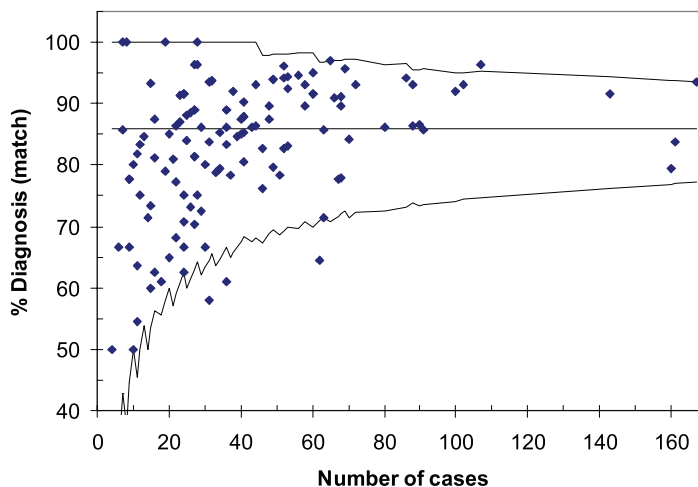


Figure 1 Diagnosis rate control chart for 131 departments.

limit indicate special-cause variation, producing high perception failure rates. There is no department below the lower control limit, and the remaining 129 departments appearing on or above the lower control limit and below the upper limit imply common-cause variation.

Fig. 5 displays on a control chart the technical failure rate for all departments. The overall technical failure rate was 4.4% (238/5454), range 0 to 30%. The 1 department on or above the upper control limit implies special-cause variation, producing high technical failure rates very unlikely to be due to chance alone with this sample size. There is no department below the lower control limit, and the remaining 130 departments appearing on or above the lower control limit and below the upper limit imply common-cause variation.

## Discussion

The overall diagnosis rate (85.9%) in this study is comparable with that of the Wessex Audit (84.6%), but the range in our study was wide (50% to 100%), as was that in Wessex (36% to 95%). To improve the diagnosis rate, we need to identify strategies for shifting the mean and reducing the tail. The control-charts methodology incorporates such guidance. The 7 departments that had particularly high diagnosis rates should be studied to learn from their apparently superior process, although the number of cases for 2 of these departments was small, i.e. less than 10. In respect of the 3 departments that had particularly low diagnosis rates, special causes need to be identified, particularly with respect to those aspects of their reporting

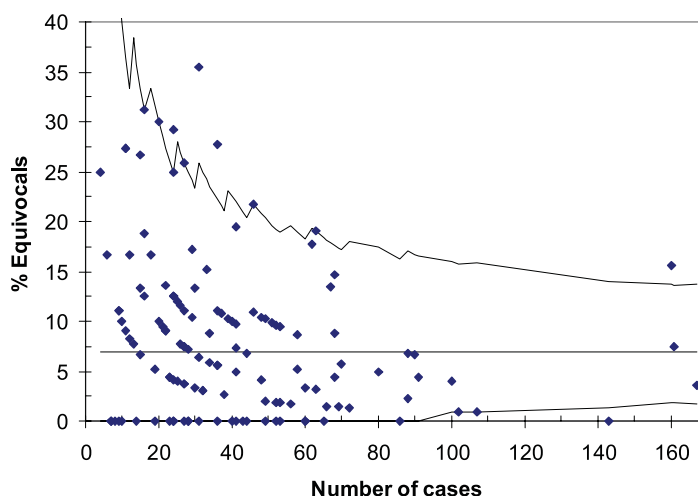


Figure 2 Equivocal rate control chart for 131 departments.

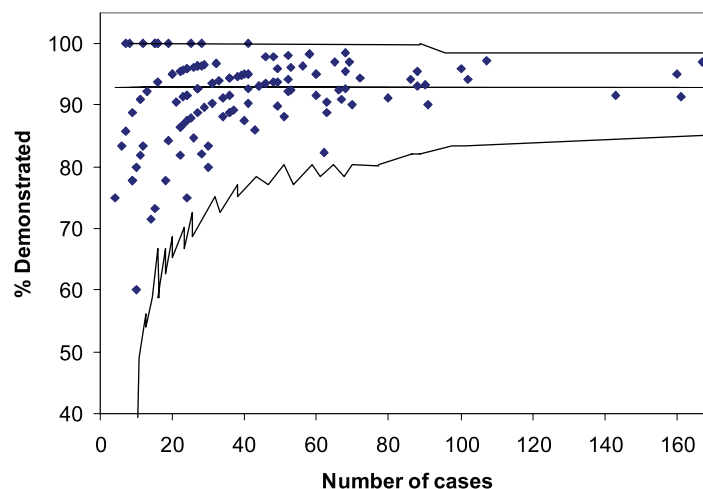


Figure 3 Demonstration rate control chart for 131 departments.

process that appear to yield such a high proportion of equivocal cases. This is confirmed by our general findings that the overall technical failure (4.4%) and perception failure (2.8%) rates were lower than the equivocal cases (6.9%), suggesting that improvement efforts should target equivocal cases in the first instance. The majority of departments (121) exhibited performance that is consistent with the limits of the current processes of reporting DBCE, and further improvement is likely to require fundamental changes to the reporting process. What sort of change? It is not possible to say at this stage, but lessons from special-cause investigation may provide some useful clues about the constraints in the present reporting process.

When investigating special-cause variation, the following generic pyramid model, adapted from industry, is recommended.<sup>12</sup> Start by checking the data, then the case-mix, the resources and the work

process, and finally check the individual practitioner. It is generally recognized that the vast majority of problems (~95%) will arise from the work system and not from individual practitioners.

The most common reason for a mismatch was an equivocal case. In our study, the proportion of equivocal cases was 6.9% (range 0 to 35.5%) compared with 8.1% (range 2% to 23%) in the Wessex Audit. To label their equivocal cases as malignant, the Wessex Audit<sup>1</sup> used criteria based on a set of predictive variables first identified and quantified by Youker and colleagues in 1968.<sup>13</sup> These criteria required, for polyps, a size > 2 cm, a broad indrawn base and an irregular surface; for strictures and obstructions, mucosal destruction and shouldering. Using these criteria, the Wessex Audit identified 26/45 (58%) of their equivocal cases as clearly malignant, a much higher score than our audit has yielded, i.e. 83/373 (22%) from internal

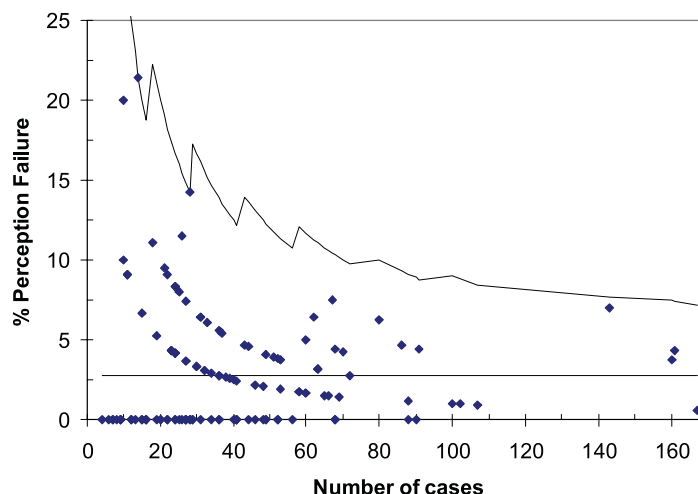


Figure 4 Perception failure control chart for 131 departments.

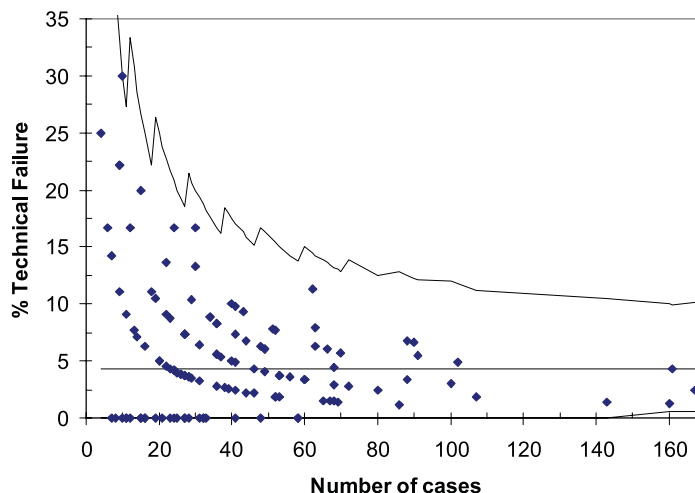


Figure 5 Technical failure control chart for 131 departments.

retrospective review using no defined criteria. Even if technique is good and perception sharp, there may still be irreducible errors of classification, and referring clinicians need to be reminded that the radiographic features used to differentiate benign from malignant are relatively non-specific.<sup>14</sup> Such cases always merit further investigation.

Part of the problem with the equivocal cases is due to the language used in the report. For example, in 1 department, some lesions were not explicitly identified as malignant; instead, the term "apple-core stricture" was used as synonymous with malignant, so other terms such as neoplastic, carcinoma and cancer were not used. Also, the wording of the report depends to some extent on the status of the referring clinician. Most radiologists now work in multidisciplinary teams. Consultant colorectal surgeons and gastroenterologists will know which lesions require further investigation, and these cases will be discussed at a multidisciplinary team meeting, so the written report may not need to be so explicit. However, if the requesting doctor is a general practitioner, then the report will specify the likely diagnosis and will recommend any further investigations. This also explains why the third target, suggesting that the report in equivocal cases should recommend further investigations in 100% of cases, was probably not appropriate. One department routinely uses a "suspicion code": 1 = normal; 2 = benign; 3 = probably benign, direct visualization recommended; 4 = probably malignant, urgent direct visualization required; 5 = malignant appearance.

The demonstration rate was 92.8% (range 60% to 100%) in our study, compared with 92.7% (range 80% to 97%) in the Wessex Audit, indicating that demonstration rates have not changed materially

during the last decade, despite changes in equipment and skill mix. This rate does not, in fact, give the true demonstration rate, as some lesions were demonstrated but not perceived by the reporting radiologist.

The perception failure rate in our study was 2.8% compared with 4.1% in the Wessex Audit. By minimizing perception failure, the demonstration rate could be improved up to 95.6% in this audit (96.8% in the Wessex Audit), i.e. the target of 95% set for this audit is achievable.

Our study is limited by being a snapshot of DBCE reporting. Longitudinal data would provide further insight and information on the performance of departments. Some departments with special-cause variation had very small sample sizes, which need to be interpreted with caution. Our study has only examined the number of true positive and false negative results (i.e., sensitivity). It has not examined, nor did it intend to examine, the false positive or true negative results (i.e., specificity). This limitation was accepted in order to keep the study relatively simple and feasible.

## Conclusion

When set against the diagnosis rates and demonstration rates reported in the Wessex Audit 1995,<sup>1</sup> it appears that a similar level of performance is currently being achieved in the NHS today, suggesting that the basic process for undertaking and reporting DCBE has remain relatively unchanged over the last few years. This indicates that improvement in the future requires fundamental changes to the process of reporting DCBE, in order to minimize the perception failure rate and

accurately to describe lesions, thus reducing the equivocal rate.

## Acknowledgements

The authors would like to thank all the radiologists of the 131 participating departments for their contribution to this audit, particularly the audit leads, who were responsible for collating and submitting the data. Without their hard work, national audits such as this would not be possible. We would also wish to thank our colleagues from the Clinical Radiology Audit Sub-Committee for their input into the preparation of this paper.

## References

1. Thomas RD, Fairhurst JJ, Frost RA. Wessex regional audit: barium enema in colo-rectal carcinoma. *Clin Radiol* 1995; **50**:647–50.
2. Stephenson G, Robertson A, Hecker R. Barium enema in the diagnosis of colonic carcinoma. *Austrian Clin Rev* 1984; **13**: 13–17.
3. Anderson N, Cook HB, Coates R. Colonoscopically detected colo-rectal cancer missed on barium enema. *Gastrointest Radiol* 1991; **16**:123–7.
4. Fort F-T. Radiographic findings in overlooked colon carcinomas: a retrospective analysis. *Acta Radiol* 1998; **29**: 331–6.
5. Kelvin FM, Gardiner R, Vas W, et al. Colo-rectal carcinoma missed on double contrast barium enema study: a problem in perception. *Am J Roentgenol* 1981; **137**:306–13.
6. Bolin S, Franzen L, Nilsson E, Sjodhal R. Carcinoma of the colon and rectum: tumours missed by radiologic examination in 61 patients. *Cancer* 1988; **61**:1999–2008.
7. Shewhart WA. *Economic control of quality of manufactured product*. New York: D Van Norstrand Company; 1931 [Reprinted by ASQC Quality Press, 1980].
8. Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; **357**:463–7.
9. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. *Br Med J* 2002; **324**: 95–8.
10. Spiegelhalter D. Funnel plots for institutional comparison. *Qual Saf Health Care* 2002; **11**:390a–1.
11. Tekkis PP, McCulloch P, Steger AC, Benjamin IS, Poloniecki JD. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *Br Med J* 2003; **326**:786–8.
12. Wheeler D. *Advanced topics in statistical process control. The power of Shewhart's charts*. Knoxville: SPC Press Inc; 1995.
13. Youker JE, Welin S, Main G. Computer analysis in the differentiation of benign and malignant polypoid lesions of the colon. *Radiology* 1968; **90**:794–6.
14. Blakeborough A, et al. Strictures of the sigmoid colon: barium enema evaluation. *Radiology* 2001; **220**:343–8.