# The clinical effectiveness of different surveillance strategies to prevent colorectal cancer in people with intermediate-grade colorectal adenomas: a retrospective cohort analysis, and psychological and economic evaluations

Wendy Atkin,[1]* Amy Brenner,[1] Jessica Martin,[1] Katherine Wooldrage,[1] Urvi Shah,[1] Fiona Lucas,[1] Paul Greliak,[1] Kevin Pack,[1] Ines Kralj-Hans,[1] Ann Thomson,[1] Sajith Perera,[1] Jill Wood,[1] Anne Miles,[2] Jane Wardle,[3] Benjamin Kearns,[4] Paul Tappenden,[4] Jonathan Myles,[5] Andrew Veitch[6] and Stephen W Duffy[5]

[1]Cancer Screening and Prevention Research Group (CSPRG), Department of Surgery and Cancer, Imperial College London, London, UK
[2]Department of Psychological Sciences, Birkbeck, University of London, London, UK
[3]Cancer Research UK Health Behaviour Centre, University College London, London, UK
[4]School of Health and Related Research (ScHARR), Health Economics and Decision Science Section, University of Sheffield, Sheffield, UK
[5]Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
[6]New Cross Hospital, Wolverhampton, UK

*Corresponding author w.atkin@imperial.ac.uk

# Scientific summary

**Effectiveness of different surveillance strategies to prevent colorectal cancer**

# Scientific summary

## Background

Colonoscopy surveillance aims to reduce colorectal cancer (CRC) incidence and mortality by removing adenomas before they become malignant and detecting cancer early. After polypectomy, CRC risk is thought to depend on the quality of the examination as well as findings at baseline, particularly the number, size and histological grade of removed adenomas, which are used to stratify patients into risk groups with different surveillance recommendations. Current UK and US surveillance guidelines divide patients with adenomas into three groups: low risk, intermediate (UK)/higher risk (US) and high risk. Both guidelines recommend 3-yearly surveillance for the intermediate-/higher-risk group. However, there remains uncertainty about the effect of surveillance on CRC risk and the optimum frequency of surveillance in the intermediate-risk (IR) group, which constitutes nearly half of all patients with adenomas.

## Objectives

The overall aim was to examine the optimum frequency of surveillance in people found to have intermediate-grade colorectal adenomas. We aimed to examine the risks and benefits to the patient with respect to prevention of CRC and the development of advanced adenomas (AAs); anxiety, morbidity and mortality; costs and cost-effectiveness; and implications for the UK NHS.

Major objectives were to assess potential heterogeneity in CRC risk according to baseline patient, polyp and procedural characteristics; to determine if there is a subgroup of IR patients who do not require surveillance, or who require one surveillance examination but no further follow-up; and to examine whether the recommended 3-year interval is too long, or could be safely extended in some patients. We also conducted an economic evaluation to estimate the incremental cost-effectiveness of alternative adenoma follow-up strategies, their impact on colonoscopy services and the total cost impact in England and Wales. In addition, a psychological assessment examined the anxiety-inducing effects of colonoscopic surveillance.

## Methods

We performed a retrospective, multicentre cohort study analysing data from two sources: a hospital data set obtained from endoscopy and pathology databases from 17 UK NHS hospitals, and a screening data set composed of three cohorts: the UK Flexible Sigmoidoscopy Screening Trial (UKFSST), the English Bowel Cancer Screening Pilot (EP) and the Kaiser Permanente Colon Cancer Prevention Program (KP). Selected hospitals had to have recorded endoscopy and pathology data electronically for at least 6 years. Eligible patients had to have had a colonoscopy at baseline and intermediate-grade adenoma(s) detected. Patients with medical conditions that increased their risk of CRC or with missing information that precluded classification of adenoma surveillance risk or surveillance intervals were excluded.

Data relating to lower gastrointestinal endoscopic procedures were extracted from hospital endoscopy databases. Pathology databases were searched for reports on colorectal lesions using Systematized Nomenclature of Medicine (SNOMED) and Systematized Nomenclature of Pathology (SNOP) codes and keywords. Corresponding endoscopy and pathology records were matched, patient identifiers removed, and the anonymised data encrypted before removal from the hospital.

A master database was created to store the endoscopy and pathology data in a standardised, structured format. Many of the data required manual interpretation and coding, so a web-based coding application was developed. Standard operating procedures (SOPs) were applied to promote uniformity of coding, and coding accuracy and data interpretation were monitored regularly.

A 'visit' was defined as one or more examinations performed in close succession (usually within 11 months) to examine the whole colon and remove detected lesions. The baseline visit included the examination at which adenomas were first diagnosed, and subsequent visits were defined as follow-up visits. Surveillance intervals were timed from the latest most complete examination of one visit to the first examination of the next visit. Patients were classified into adenoma surveillance risk groups using the UK guideline:

- *Low risk* One or two small (< 10 mm) adenomas, no large ($\geq$ 10 mm) adenomas.
- *Intermediate risk* Three or four small adenomas, none of which is large; or one or two adenomas, at least one of which is large.
- *High risk* Five or more adenomas (any size); or three or more adenomas, at least one of which is large.

The primary outcomes were CRC incidence after baseline and first follow-up, and AA and CRC detected at first and second follow-up visits (FUV1 and FUV2). An AA was defined as an adenoma of $\geq$ 10 mm, or with villous or tubulovillous histology, or high-grade dysplasia (HGD). In some analyses, AAs and CRCs that were seen at a prior visit were excluded, as they were considered a surrogate for polypectomy site surveillance, which could confound the analyses. CRC was ascertained using pathological data recorded on the study database and from national sources. The main exposures of interest were number of surveillance visits and length of surveillance interval to first or second follow-up. Patient, procedural and polyp characteristics were assessed as a priori risk factors and confounders.

Logistic regression was used to model the association of surveillance interval length and secondary risk factors with findings at first and second follow-ups. Survival analysis was used to assess the effect of surveillance on risk of CRC after baseline, and after first follow-up. All time-to-event data were censored at first CRC diagnosis, death, emigration or end of follow-up. Time at risk started from the last most complete colonoscopy in baseline or the first procedure in FUV1, and if CRC was diagnosed at a follow-up visit then that follow-up visit was not included, as it did not offer any protection against CRC. 'One minus the Kaplan–Meier estimator of the survival function' was used to illustrate the time to cancer diagnosis and to estimate the cumulative risk of cancer with 95% confidence intervals (CIs) at 3, 5 and 10 years. Independent predictors were identified in multivariable models, using backward stepwise selection with a *p*-value of < 0.05 in the likelihood ratio test to determine the retention of variables in the final logistic and Cox regression models.

In the psychological analysis, participants were men and women aged 55–64 years, at average risk of CRC, invited for screening in the UKFSST (*n* = 35,891). Pre- and post-screening questionnaires were used to assess negative and positive emotional consequences in participants with no polyps detected by flexible sigmoidoscopy (FS), low-risk polyps removed at FS, high-risk polyps referred for colonoscopy or high-risk polyps recommended for surveillance. Between-group analyses of variance (ANOVAs) were used to test for differences between screening outcome groups.

The health-economic analysis used IR patients in the hospital data set. Thirteen alternative surveillance strategies were evaluated, taking into account a range of alternative surveillance intervals and the presence/absence of a cut-off for eligibility, based on patient age. A cost–utility analysis was undertaken, whereby the primary health-economic outcome was defined in terms of the incremental cost per quality-adjusted life-year (QALY) gained.

## Results

### Hospital data set

There were 11,944 eligible IR patients, 55.5% of whom were male, and the median age was 66.7 years [interquartile range (IQR) 58.4–74.0 years]; 4608 had at least one follow-up and 1635 had two or more. A total of 168 CRCs developed during 81,442 person-years (pys) of observation time after baseline (206 per 100,000 pys, 95% CI 177 to 240 pys). A single follow-up visit was associated with a 49% lower CRC incidence than no surveillance [hazard ratio (HR) 0.51, 95% CI 0.34 to 0.77], after adjusting for covariates. Having two or more surveillance examinations conferred a further 19% reduction in CRC risk (HR 0.32, 95% CI 0.17 to 0.61). Lower-intermediate-risk (LIR) and higher-intermediate-risk (HIR) subgroups were identified based on polyp and procedural risk factors for CRC identified in the Cox models. The HIR subgroup included patients with any of the following baseline characteristics: an adenoma of $\geq$ 20 mm or with HGD, proximal polyps, no complete colonoscopy or poor bowel preparation. All other patients were assigned to the LIR subgroup. The subgroups comprised 9265 HIR (77.6%) patients and 2679 LIR (22.4%) patients. CRC risk was 69% lower in the LIR than in the HIR subgroup (HR 0.31, 95% CI 0.18 to 0.55; $p < 0.0001$). In the HIR subgroup, one follow-up conferred a 50% reduction in risk (HR 0.50, 95% CI 0.34 to 0.76) and two or more follow-ups a 64% reduction in risk (HR 0.36, 95% CI 0.20 to 0.62). In the LIR subgroup a single follow-up conferred a 38% reduction in CRC risk. However, this result was non-significant; thus, the benefit of surveillance in the LIR subgroup remained unclear (HR 0.62, 95% CI 0.16 to 2.43; $p = 0.4700$).

Of the 4608 patients who attended FUV1, 451 (9.8%) had AA and 52 (1.1%) had CRC detected. An 18% increased odds of new AA and a 32% increased odds of CRC were seen per year increase in interval to first follow-up [AA: odds ratio (OR) 1.18, 95% CI 1.12 to 1.24; CRC: OR 1.32, 95% CI 1.20 to 1.46]. Of the 1635 patients who attended the second follow-up, 146 (8.9%) had new AA and nine (0.6%) had new CRC detected. At FUV2, a 22% increased odds of new advanced neoplasia (AN: AA or CRC) was seen per year increase in interval (OR 1.22, 1.09 to 1.36); the association was significant only after adjusting for confounding factors. At FUV1, there was a significant association between interval and new AN in both the HIR ($p < 0.0001$) and LIR ($p = 0.0433$) subgroups. At FUV2, there was an association between interval and new AN in the HIR subgroup ($p = 0.0191$), but not in the LIR subgroup ($p = 0.4573$). Detection rates of new AA and CRC at follow-up remained at < 10% and < 1%, respectively, with an interval of < 3 years in all IR subgroups except in patients with a poor examination.

### Screening data set

In the pooled data set there were 2352 individuals: 67.8% were male, and the median age was 61.5 years (IQR 58.0–65.0); 1828 had at least one follow-up and 1011 had two or more. A total of 32 CRCs developed during 25,745 pys of observation time after baseline (124 per 100,000 pys, 95% CI 88 to 176 pys). One follow-up conferred a significant 73% reduction in the incidence of CRC (HR 0.27, 95% CI 0.10 to 0.71) after adjusting for risk factors identified in the hospital data set. Additional surveillance did not appear to provide further protection. Participants were divided into HIR and LIR subgroups using the definition derived from risk factors for CRC in the hospital data set. In the HIR subgroup, a single surveillance visit conferred a 72% reduction in CRC risk (HR 0.28, 95% CI 0.09 to 0.92; $p = 0.0508$), and in the LIR subgroup there was a non-significant reduction in risk (HR 0.25, 95% CI 0.05 to 1.30; $p = 0.2084$).

In the 1828 (77.7%) screening participants who attended surveillance, AA and CRC were found at the first follow-up in 4.2% and 0.7%, respectively. When models including the same set of predictors for findings at first follow-up in the hospital data set were fitted to the screening data set, no association was found between increasing interval to first follow-up and detection of AA or CRC (AA: OR 0.98, CI 0.86 to 1.12; $p = 0.72$; CRC: OR 1.12, CI 0.84 to 1.48; $p = 0.46$). Similarly, no significant difference in risk of AA or CRC at first follow-up was found between the LIR and HIR subgroups defined using long-term CRC risk in the hospital data set.

Bowel cancer worry differed by group ($F_{3,31904} = 16.3$; $p < 0.001$): there was more worry in the surveillance group than in the no-polyps group, and no significant differences between the surveillance group and the other two groups with polyps. Although general psychological distress differed by outcome group ($F_{3,32055} = 2.66$; $p < 0.05$), the surveillance group reported less distress than the groups with no polyps or lower-risk polyps. There were significant differences across the groups in reported emotional consequences of screening ($F_{3,31971} = 9.37$; $p < 0.001$), with the surveillance group reporting higher positive consequences of screening than all of the other groups. They also reported more reassurance than the lower-risk group, although reassurance scores did not differ from the two remaining outcome groups.

### Health-economic analysis.
Of the 13 surveillance options considered in the exploratory economic evaluation, 3-yearly ongoing colonoscopic surveillance with no age limit produced the greatest expected health gain. The incremental cost-effectiveness ratio (ICER) for this option (compared with an age cut-off of 75 years) was expected to be < £3000 per QALY gained. The probabilistic sensitivity analysis indicated that, assuming a willingness-to-pay threshold of £20,000 per QALY gained, the probability that this option would produce the greatest expected net benefit was approximately 1.0. When compared marginally against the no-surveillance strategy, all surveillance options either dominated or had an ICER that was < £1000 per QALY gained.

## Conclusions

Our results from both the hospital and screening data sets provide strong evidence that a single follow-up offers substantial benefit to all IR patients by lowering their future risk of CRC. The benefits of a second surveillance were more modest. However, we identified a higher-risk subgroup that significantly benefited from additional surveillance and, conversely, a lower-risk subgroup for whom additional surveillance may not be necessary.

If CRC risk is considered high enough to warrant surveillance in a subgroup, our data suggest that it should be done at 3 years, unless the prior examination is of poor quality, in which case an earlier examination should be considered. Surveillance should not be delayed as rates of interval cancers increase. The economic analysis confirmed that 3-yearly surveillance was the optimum strategy. Further gains might be expected from an identified lower-risk subgroup for which a single surveillance visit might suffice. The detection of polyps was associated with increased bowel cancer worry, but surveillance itself was associated with improved psychological well-being.

Future studies are needed to further validate findings in this study and confirm whether or not some IR patients may not need colonoscopic surveillance.

## Study registration

This study is registered as ISRCTN15213649.

## Funding

# Health Technology Assessment

HTA/HTA TAR

## Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

## HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

## This report